

TARTU ÜLIKOOL
Majandusteaduskond

Margus Matkur

**PROGNOOSIMINE GOOGLE OTSINGU MÄRKŠÕNADE
SAGEDUSE ABIL AEGRIDADE MEETODITEGA**

Magistritöö ärijuhtimise magistrikraadi taotlemiseks
ettevõtluse ja tehnoloogia erialal

Juhendaja: lektor Andres Võrk

Tartu 2017

Soovitan suunata kaitsmisele

(juhendaja nimi)

Kaitsmisele lubatud “ “..... 2017. a

Olen koostanud töö iseseisvalt. Kõik töö koostamisel kasutatud teiste autorite tööd, põhimõttelised seisukohad, kirjandusallikatest ja mujalt pärinevad andmed on viidatud.

.....

(töö autori nimi)

SISUKORD

SISSEJUHATUS	4
1. INTERNETI SUURANDMETE KASUTAMINE PROGNOOSIMISEL.....	6
1.1. Prognoosimudelite kasutamise valdkonnad ja sihtrühmad	6
1.2. Ülevaade varasematest uurimistöödest	9
2. ANDMED JA METOODIKA.....	14
2.1. Uuritavate näitajate statistilised andmed.....	14
2.2. Uuritavate näitajate Interneti-otsingute andmed ja Google indeksi koostamise metoodika	16
2.3. Prognoosimise meetodid	22
3. PROGNOOSIMUDELITE HINDAMINE JA TULEMUSED.....	27
3.1. Aegridade graafiline analüüs.....	27
3.2. Prognoosimudelid.....	32
3.3. Mineviku prognoosimine	41
3.4. Tuleviku prognoosimine	47
3.5. Algoritm Google otsingu märksõnade sageduse andmete kaasamiseks prognoosimudelitesse.	54
KOKKUVÕTE.....	56
VIIDATUD ALLIKAD	59
LISAD	66
Lisa 1. Statistika metoodikad	66
Lisa 2. Prognoosimudelite valemid	74
Lisa 3. Joonas Tuhkuri tööde prognoosid.....	79
SUMMARY	81

SISSEJUHATUS

Eesti statistikaameti andmetel on Eestis arvuti ja Interneti kasutajaid 2015 aasta andmetel 16-74 aastaste elanike seas üle 852 tuhande, mis moodustab 88% vastava rühma koguarvust (Statistikaamet 2016). Maailmas on arvuti ja Interneti kasutajaid üle 3 miljardi (InternetLiveStats.com 2016). 92% Interneti kasutajatest on kasutanud otsingumootoreid info otsimiseks Internetist (GO-Gulf 2012). Tehnoloogiaettevõtte Google Inc. otsingumootor www.google.com omab 89% globaalsest turuosast otsingumootorite seas (Statista Inc. 2016) ning www.google.ee on Eestis kõige kasutatavam internetileht (Alexa Internet Inc. 2016). Google otsingumootorit kasutades tehakse maailmas üle 57 tuhande päringu sekundis, see teeb üle 1,7 triljoni päringu aastas (InternetLiveStats.com 2016). Otsingute käigus talletuvad suured andmemahud ehk suurandmed - *Big Data*. Tekib küsimus, kas sellest suurest andmemassiivist on võimalik kasulikku lisateavet välja noppida nii ettevõtete juhtidel kui ka majanduspoliitika kujundajatel?

Käesoleva magistritöö eesmärk on uurida, kas Internetis teostatud otsingute andmetest on abi majandusprotsesside prognoosimisel nelja Eesti majandusprotsessi näitaja näitel. Kiirem ja täpsem majandusnäitajate prognoos võib aidata otsustajaid nii ettevõtte tasandil, näiteks finantsasutustes, tööstusettevõtetes, kui ka riigi tasandil, otsustamaks, kas majandust on tabamas stagnatsioon või milline sektor on majanduses ülekuumenemas. Et ametlik statistika majandusnäitajate kohta laekub sageli mitme kuulise viitajaga, kuid inimeste poolt tehtud otsinguandmed on kättesaadavad praktiliselt reaalajas, siis võib otsinguandmete kasutamine võimaldada relevantsema informatsiooni kiiremat kasutamist prognooside tegemisel. Käesoleva magistritöö eesmärk ei ole leida parim prognoosimudel vaid hinnata Google otsingusõnade statistika võimekust majandusprotsesside prognooside parendamiseks.

Töö ülesanded on.

1. Anda ülevaade valdkonna varasematest uurimistöödest.
2. Tutvustada konkreetse majandusprotsessi näitaja statistilisi andmeid ja koguda Google Trends vaatlusandmed ning nende baasil koostada Google indeks (GI).
3. Teostada iga näitaja kohta Google indeksi ja majandusprotsessi graafiline analüüs.
4. Konstrueerida prognoosimudelid ja teostada mineviku prognoosimine ning arvutada prognoosi ja tegelike väärtuste vaheline viga.
5. Pakkuda välja tuleviku prognoosid nelja Eesti majandusprotsessi kohta.
6. Formuleerida algoritm Google otsingute andmete kaasamiseks prognoosimudelitesse.

Töö on teemast lähtuvalt empiirilise suunitlusega ja koosneb kolmest peatükist. Esimeses peatükis käsitletakse teemaga seotud valdkondi ja teemaga seotud valdkondadest huvitatud sihtrühmi ning antakse ülevaade varasemalt teostatud töödest. Teises peatükis esitatakse nelja vaatluse all oleva näitaja statistilised andmed: töötuse määr, majutatute arv, jaemüük posti või interneti teel ning kinnisvara ostu-müügitehingute arv. Valitud valdkonnad on ühest küljest seotud majanduse üldise arenguga, teisest küljest eksisteerib valdkondade vastu elanikkonna otsene huvi ehk Interneti-otsingud. Seejärel selgitatakse Google Trends rakenduse toimimist, kirjeldatakse vaatlustulemuse kogumist ja moodustatakse iga vaatluse all oleva näitaja jaoks Google indeks (GI). Antakse ülevaade võimalikest prognoosimise lähenemistest. Kolmandas peatükis teostatakse Google indeksi ja majandusprotsesside näitajate graafilised analüüsid ja selgitatakse prognoosimodelite ülesehitust ning koostatakse prognoosimudelid. Seejärel esitatakse mineviku prognoosimise arvutustulemusi, mille võrdlusest tegelike, minevikust teadaolevate andmetega peab selguma, kas Google otsingute andmete kaasamine prognoosimudelitesse on põhjendatud. Lisaks pakutakse välja tuleviku prognoosid iga vaatluse all oleva näitaja kohta. Töö tulemusena formuleeritava algoritmi põhjal on võimalik üles ehitada tarkvara rakendus, mis reaajas tegeleb majandusnäitajate prognoosimisega. Antud töös kasutatakse prognoosimismudelite loomisel lähtekohana Hyunyoung Choi, Hal Varian (2009a, 2009b) ja Joonas Tuhkuri (2014 ja 2016) töid.

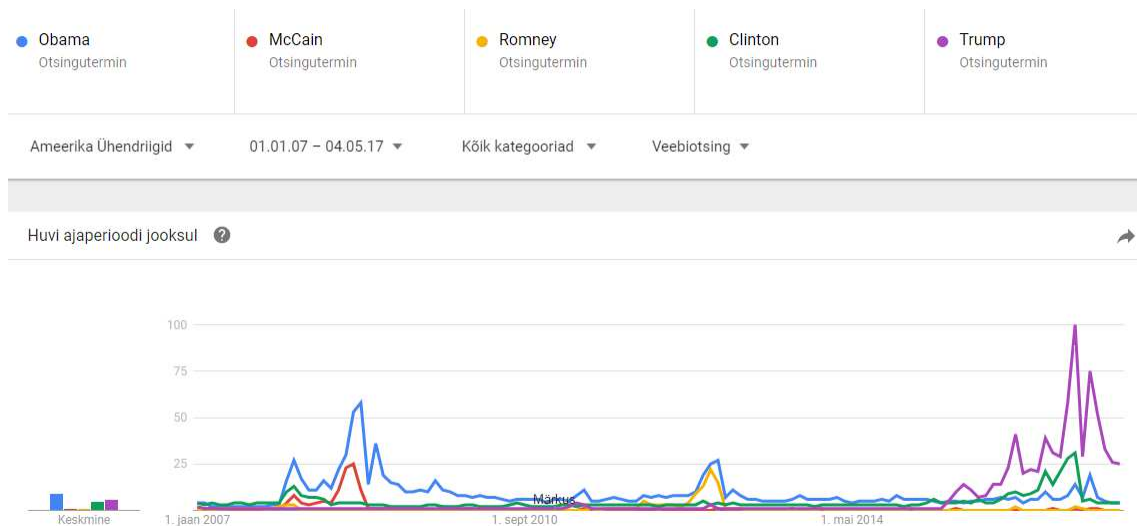
1. INTERNETI SUURANDMETE KASUTAMINE PROGNOOSIMISEL

1.1. Prognoosimudelite kasutamise valdkonnad ja sihtrühmad

Interneti kasutamisel inimeste poolt talletatakse erinevates Interneti rakendustes ja andmebaasides määratu hulk kõikvõimalike andmeid selle kohta, kes, mida ja millal postitas, päris, sisestas, külastas jne. Tavaliselt ei ole vastavad andmed avalikud ega ka koheselt kasutatavad. Näiteks koguvad sideoperaatorid ja Internetis erinevaid teenuseid pakkuvad ettevõtted vastavaid andmelogisid, kuid nendest väljavõtete tegemine ei ole reeglina tavakasutajatele kättesaadavad. Neid andmeid saavad kasutada kas ettevõtte ise või riiklikud jõustruktuurid. Samas esineb näiteid, kus Internetis mingi teenuse pakkuja avalikustab vähemalt osa tema teenusega seotud kasutusstatistikat. Näiteks avaldavad www.google.com (Google Trends) ja www.yandex.com (wordstat.yandex.com) otsingumootorid statistikat otsingute märksõnade esinemise sageduse kohta, twitter.com (Twitter Analytics) vahendusel saab jälgida säutsude (*tweet*) aktiivsust, Facebook.com lehelt saab jälgida mingi teema järgijate ja meeldivuste (*follow, like*) arvu. Kuna tegemist on globaalsete teenuste pakkujatega ehk nende kasutajaid on väga palju ja erinevatest geograafilistest piirkondadest ning andmed on avalikud, siis võib tekkida mõte analüüsida pakutavaid andmeid ja nende baasil püüda mingit lisateavet tekitada. Hea näide siinkohal on eesti päritolu teadlase Kalev Hannes Leetaru tööd, uurimaks uudiste tonaalsuse seost poliitiliste protsessidega *Culturomics 2.0* (Leetaru 2011) või projekt GDELT (www.gdeltproject.org) ehk maailma uudiste ning nende seoste andmebaasi loomine. Avalike andmeid on võimalik kasutada, et uurida ühiskonna sotsiaalseid, majanduslikke, tervishoiu, poliitilisi jne. valdkondi. Lisaks akadeemilisele kogukonnale võivad eelnimetatud valdkondade uurimisest huvitatud olla näiteks riigi- ja finantsasutused ning meediaettevõtted. Piiratud ligipääsuga andmed võivad olla uurimisobjektiks organisatsioonisiselt või pakkuda spetsiifilist huvi riikide jõuametkondadele.

Vaatamata sellele, kas andmed on avalikud või piiratud ligipääsuga, on samad andmetöötluste ja analüüsi meetodid kasutatavad mõlemal juhul.

Interneti-otsingute märksõnade sageduse andmete analüüsi üks lihtne näide on tõdemus, et alates aastast 2004, kui Google Trend alustas Google otsingute märksõnade sageduse andmete avaldamist, on iga Ameerika Ühendriikide presidendivalimise võitnud see kandidaat, kes on Google otsingutes oma vastaskandidaadist valimiste eel populaarsem. Eriti kurioosne oli olukord 2016. aasta presidendivalimiste aegu, kui enamus tavaprognosidest pakkus valimiste võitjaks Hillary Clintonit, kuigi Google otsingutes edastas Donald Trump Clintonit märkimisväärselt. 2016 aasta Ameerika Ühendriikide presidendivalimised võitis Donald Trump. Tõdemusest, et siia maani on alati presidendivalimised võitnud kandidaat, kes on oma vastaskandidaadist Google otsingutes populaarsem, saab näiteks teha järeldusi järgmiste presidendivalimiste kandidaatide ülesseadmisel või valimiste jälgimisel.



Joonis 1. Ameerika Ühendriikide presidendivalimiste kandidaadid Google otsingutes (Allikas: Google Trends 2017)

Twitteri andmetele toetudes on uuritud näiteks gripilaadsete viirushaiguste levimise trende (Achrekar, Gandhe, Lazarus, Yu, Liu 2011), poliitilisi valimisi (Jahanbakhsh, Moon 2014), kuritegude toimumist (Bendler, Brandt, Wagner, Neumann 2014), väärtpaberituru indikaatorite prognoosimist (Zhang, Fuehres, Gloor 2010), kinofilmide läbimüüki (Asur, Huberman 2010) ning tarbijate kindlustunnet (O'Connor, Balasubramanyan, Routledge, Smith 2010). Facebooki andmetele toetudes on uuritud

muuhulgas üldist rahulolu (Collins, Sun, Kosinski, Stillwell, Markuzon 2015). Erakorralise meditsiini osakonda (EMO) pöördumiste prognoosimist on uuritud kasutades näiteks tervishoiuasutuse kodulehekülje külastatavuse statistikat (Ekström, Kurland, Farrokhnia, Castrén, Nordberg 2013). Eelpool nimetatud tööd ei ole oma valdkonnas reeglina ainukesed.

Kõikidest Interneti kasutamisega seotud algandmete analüüsimise töödest kumab läbi, et mõtet on tegeleda nende valdkondadega, millel on piisavalt suur ühiskondlik kõlapind ehk kasutajaid, kes vastavaid andmeid aja kulgedes genereeriks, oleks võimalikult palju. See tagab ka algandmete piisava hulga analüüside teostamiseks. Samas on sihtrühm, keda võiksid ühiskonna, sealhulgas majanduslikud protsessid huvitada, mitte ainult eelpool mainitud organisatsioonid, vaid ka üksikisikud. Näiteks võib mõni üksikinvestor olla huvitatud ühiskondliku huvi seiramisest, et selle abil teha õigeid investeerimisotsuseid. Nähtavasti eelpoolmainitud huvist lähtuvalt on tehtud palju töid börsil kauplemise valdkonna kohta. Autorile teadaolevalt on riiklikud ametkonnad nagu Eurostat ning finantsasutused huvitatud majanduse tervisliku olukorra kiirhinnangute andmisest (*nowcasting*) kui ametliku statistika andmete avaldamine võtab liiga kaua aega (Baldacci, Wirthmann, Kovachev, Golmajer 2016). Siinkohal ongi võimalik kasutada prognooside tegemisel enam mitte nii ajakohastele ametliku statistika andmetele lisaks veel Interneti poolt pakutavaid ajakohaseid suurandmeid. Samas on võimalik kasutada analoogset lähenemist ka mitteavalike andmete prognoosimiseks, mida pakuvad muud tehnoloogilised lahendused. Näitena võib tuua intelligentsete tehnovõrkude (elektrivõrk, andmesidevõrgud, automaatjuhtimissüsteemid jne.) poolt pakutavad andmed. Ühena rakendustest võib välja tuua mobiilsidevõrgud, kus on võimalik kaasata ajaloolistele arveldusandmetele lisaks reaalaajaandmeid võrgus toimuva kohta selleks, et ennetada või õigeaegselt avastada petturluse (*fraud*) juhtumeid. Teisisõnu jälgitakse võrgus toimuvat reaajas ning selle tulemusena avastatakse võimalikud petturluse juhtumid koheselt mitte aga peale arveldusperioodi lõppemist ehk tagantjärele raamatupidamisraporteid analüüsid. See võimaldab juhtumi avaldumisel kohest reageerimist ja suuremate kahjude ärahoidmist (*revenue assurance*).

Käesoleva töö valdkondade valiku kriteeriumiteks on asjaolud, et valdkond oleks seotud majanduse üldise arenguga ning, et nendega seotud märksõnade kohta oleks olemas

analüüsi teostamiseks piisavalt Interneti-otsingute andmeid. Teisisõnu, uuritava majandusvaldkonnakonna vastu peab eksisteerima laiem elanikkonna kõrgendatud huvi. Loomulikult peab valitud näitajate kohta olema olemas ka ajaloolised statistilised andmed. Töös kaasatakse prognoosimise läbiviimiseks lisaks uuritava majandusliku näitaja ajalooliste väärtustele ka tehnoloogiaettevõtte Google Inc. poolt pakutavaid valdkonnaga seotud väliseid andmeid. Tulemuseks saadakse hinnang, millised võiksid olla näitajate arväärtused käesolevaks hetkeks (*nowcast*) või siis tulevikus (*forecast*). Kõik töös kasutatavad andmed on avalikud.

1.2. Ülevaade varasematest uurimistöödest

Internetis teostatud otsingute andmete alusel ühiskonna protsesside prognoosimise uurimisel oli teedrajav Ginsberg, Mohebbi, Patel, Brammer, Smolinski ja Brilliant (2009) töö, kus suudeti grupi gripilaadsete viirushaiguste Google otsingu märksõnade esinemise sageduse baasil prognoosida gripilaadsete viirushaigustega seoses arstile pöördunute arvu lähitulevikus. Selleks kasutas Ginsbergi töörühm Google Inc. poolt salvestatud andmeid, et sisuliselt reaajas seirata gripilaadsete viirushaiguste levimist asukohtades, kus Interneti kasutamine oli laialt levinud. Sellele järgnesid Choi ja Varian (2009a; 2009b) tööd, kus näidati, et Interneti-otsingute andmete abil on võimalik parendada jaemüügi, mootorsõidukite ja nende osade jaemüügi, kinnisvaraturu, Hong-Kongi reisimise ning esmakordsete töötushüvitistaotluste arvu prognoosimist. Kui senised tööd olid olulisel määral Ameerika Ühendriikide kesksed, siis selle korvasid veel 2009 aastal Askitas, Zimmermann (2009) (Saksamaa – töötuse määr), Suhoy (2009) (Iisrael - töøjõuturg, koduelektroonika, kinnisvara, iluteenused, jaekaubandus, turism), D'Amuri (2009) (Itaalia - töötuse määr). Kõik nad leidsid, et Google otsingumootoris otsitavate märksõnade analüüsil on tugev potentsiaal, et seda kasutada majanduslike protsesside prognoosimisel. Järgnevatel aastatel on lisandunud kümneid töid erinevate maade ja majandusnäitajate kohta alates inflatsiooni prognoosimisest (Guzman 2011) lõpetades börsil kauplemise otsuste valikul, toetudes Google otsingumootoris teostatud päringute statistikale (Preis, Moat, Stanley 2013). Autorile teadaolevalt on Eesti lähiriikide kohta teostanud Interneti-otsingute alusel majanduslike protsesside prognoosimise uuringuid Lindberg (2011) (Rootsi – jaemüük) ja Tuhkuri (2014) (Soome – töötuse määr).

Tabel 1. Lühiülevaade varasematest Google Trends andmeid kasutanud töödest

Autorid	Valdkond	Sihtriik	Tulemus
Ginsberg, Mohebbi, Patel, Brammer, Smolinski, Brilliant (2009)	Gripilaadsete viirushaigustega seoses arstile pöördunute arv	Ameerika Ühendriigid	2007-2008 aasta gripilaadsete viirushaiguste hooaja jooksul saatis töörühm oma prognoositavad andmed <i>US Centers for Disease Control and Prevention</i> (CDC) organisatsiooni, et hinnata nende ajastust ja täpsust. Töörühm suutis prognoosida järjepidevalt andmeid 1-2 nädalat ettepoole CDC enda raportitest.
Choi, Varian (2009a)	Jaemüük, mootorsõidukite ja nende osade jaemüük, kinnisvara, reisimine Hong-Kongi	Ameerika Ühendriigid, kogu maailm	Leitakse, et mudelid, mis sisaldavad Google Trend otsingu muutujate andmeid, edestavad mudeleid, mis ei sisalda Google Trend otsingu muutujate andmeid. Osade mudelite parendus on kõigest mõned protsendid, kuid osade mudelite parendus on märkimisväärne (kuni 18%).
Askitas, Zimmermann (2009)	Töötuse määr	Saksamaa	Töös näidatakse tugevat korrelatsiooni märksõnade otsingu andmete ja töötuse määra vahel. Nenditakse, et meetodil on tugev potentsiaal töötuse määra prognoosimiseks ja seega ka edasisteks uuringuteks.
Suhoy (2009)	Tööstus, koduelektronika, kinnisvara, iluteenused, jaekaubandus, turism	Iisrael	Iisraeli andmed toetavad hüpoteesi, et Google päringu andmed võivad olla abiks iseloomustamiseks riigi hetkelist majanduse seisust. Kõige paremini prognoositav näide on tööstus.
Choi, Varian (2009b)	Esmakordsete töötushüvitistaotluste arv (<i>Initial Claims for Unemployment Benefits</i>)	Ameerika Ühendriigid	Näidatakse, kuidas Google Trends andmed aitavad prognoosida esmakordsete töötushüvitistaotluste arvu.
D'Amuri (2009)	Töötuse määr.	Itaalia	Google indeks töötab üsna hästi (<i>“performs fairly well”</i>) töötuse määra prognoosimisel ja ületab teisi juhtivaid indikaatoreid nagu töötuse ootuse uuringud ja tööstustoodangu indeks.
Vosen, Schmidt (2009)	Eratarbimine.	Ameerika Ühendriigid	Töös näidatakse, et Google Trends on väga lubav uus andmete allikas, et prognoosida eratarbimise käitumist.

Kulkarni, Haynes, Stough, Paelinck (2009)	Kinnisvara	Ameerika Ühendriigid	Õigesti valitud Interneti-otsingute märksõnade valimise puhul on võimalik leida nende mõju vastavale <i>Case-Shiller</i> ’i kinnisvara indeksile.
Kholodilin, Podstawski, Siliverstovs (2010)	Eratarbimine	Ameerika Ühendriigid	Mudelite võrdlemisel leitakse statistiline tõestus, et mudelid, mis kasutavad Google indikaatoreid parendavad prognoosimise täpsust võrreldes mudelitega, mis ei kasuta Google indikaatoreid.
Moen, Anvik, Gjellstad (2010)	Töötuse määr	Norra	Töös järeldatakse, et Google otsingu tulemused sisaldavad informatsiooni, mida on võimalik ära kasutada töötuse määra lühiajalisel prognoosimisel.
Bughin, (2011)	Jaekaubandus, töötus	Belgia	Töö kinnitab eelmiste samalaadsete tööde järeldusi. Töös kasutatud näidete lähiaja prognoosimise (<i>nowcasting</i>) võimekus on statistiliselt märkimisväärne majanduslike trendide tuvastamiseks.
McLaren, Shanbhogue (2011)	Tööjõuturg, kinnisvaraturg	Ühend- kuningriik	Vaatamata Google poolt pakutavate andmete piirangutele (tegelike päringute arv, tagantjäre andmete muutumine) järeldatakse, et need andmed on siiski kasutatavad prognooside tegemiseks.
Guzman (2011)	Inflatsioon	Ameerika Ühendriigid	GISI (<i>Google Inflation Search Index</i>) edastab PCE (<i>personal consumption expenditure</i>) indeksit 12 kuuga ja omab madalamat prognoosimise viga võrreldes kõikide indikaatoritega.
Pan, Wu, Song (2012)	Majutatute arv	Charleston, Ameerika Ühendriigid	Analüüs näitab, et viie Google Trends märksõna andmete lisamine mudelitesse tõstis prognoosimise täpsust oluliselt.
Chadwick, Sengul (2012)	Mittepõllumajanduslik töötuse määr	Türgi	Töötuse määra prognoosimise mudelites kasutatakse Google indeksit GI. Analüüsitakse erinevaid mudeleid. Loodud mudelid ennustavad paremini (kuni 3 kuud ette) võrreldes lihtsalt antud hetkel teadaoleva (ajaloolise) töötuse määra aluseks võtva mudeliga.
Chmyznikov, Galvanauskas (2013)	Töötuse määr, autode müük, elamispindade müük, inflatsioon	Eesti, Läti Leedu	Vastavad Google otsingute andmed sarnanevad Baltimaade vastavate majandusnäitajate andmetega. Sellest tuleneb, et Google otsingute andmed on kasulikud töötuse määra, autode müügi, elamispindade müügi ja inflatsiooni lühiajalise prognoosimisel ning andmeid on võimalik kasutada majandusaktiivsuse lisaindikaatorina.

Preis, Moat, Stanley (2013)	Kauplemine aktsiaturgudel	Ameerika Ühendriigid, kogu maailm	Finantsvaldkonnaga seotud märksõnade otsimine Google otsingumootoris kasvab enne aktsiaturgude langust. Korrelatsioon on selgem Ameerika Ühendriikide puhul. Kui investeerimisstrateegiate võrduses tootis „ <i>buy and hold</i> ” kasumit 16%, siis „ <i>Google Trends strategy</i> “ 326%.
Tuhkuri (2014)	Töötuse määr	Soome	Mudelid, mis sisaldavad Google Trends poolt pakutavaid andmeid, pakuvad täpsemaid prognoosimise tulemusi, kui samad mudelid, mis ei sisalda Google Trends poolt pakutavaid andmeid.
Wu, Brynjolfsson (2014)	Kinnisvara	Ameerika Ühendriigid	Interneti-otsingute andmete analüüsimine, kasutades suhteliselt lihtsaid mudeleid, võib viia palju täpsemate prognoosideni, kui see oli võimalik eelnevalt. Google Trensi poolt kogutud mikroandmed võivad osutada üheks oluliseks tööriistaks, mille abil tarbijad, ettevõtted ja ametnikud saavad teostada prognoose, et teha tõhusaid otsuseid.
Pavlicek, Kristoufek (2015)	Töötuse määr	Tšehhimaal, Ungari, Poola, Slovakkia	Ülesehitatud prognoosimise mudelid järgivad töötuse määra ajaloolist muutumise dünaamikat. Eriti hästi tuleb see esile Tšehhi ja Ungari näite varal. Poola ja Slovakkia puhul on seos väiksem. Autorid selgitavad seda Interneti madalama kasutuslevikuga ja tõsiasjaga, et Poola ja Slovakkia elanikud on rohkem nõus töö tegemisel reisima väljapoole oma riiki kui Tšehhi või Ungari elanikud.
Tuhkuri (2016)	Töötuse määr	Ameerika Ühendriigid	Mudelid, mis sisaldavad Google Trends poolt pakutavaid andmeid, pakuvad keskmiselt täpsemaid prognoosimise tulemusi, kui samad mudelid, mis ei sisalda Google Trends poolt pakutavaid andmeid.

Allikas: (viidatud autorite tööd, täpsemalt viidatud allikate osas)

Eesti kohta on uuringu avaldanud Chmyznikov, Galvanauskas (2013), kus vaatluse all oli Eesti, Läti ja Leedu töötuse määr, autode müük, elamispindade müük ja inflatsioon. Tabelis 1 on ära toodud lühiülevaade varasematest Google Trends andmeid kasutanud töödest. Nagu tabelist näha, on tööde valdkonnad ja sihtriigid erinevad. Valdkondadest on esindatud muuhulgas meditsiin, jaekaubandus, majutus, üldine tööpuudus ja kinnisvaraturg. Üldjuhul kasutatakse töödes kuu detailsusega andmeid (töötus, jaekaubandus jne.) kuid näiteks meditsiini valdkonna töös viiakse prognoosimist läbi nädalase detailsusega (Ginsberg, Mohebbi, Patel, Brammer, Smolinski ja Brilliant 2009). Üldiselt leitakse varasemates Google Trends andmeid kasutanud uurimistöödes, et Google otsingute märksõnade sageduse andmed sisaldavad informatsiooni, mida on võimalik ära kasutada prognooside parendamisel. Samas tuleb siiski välja tuua, et tehtud tööde fookus, kasutatud meetodid ja detailsuse aste võivad olla erinevad. *Big Data* andmemahutuste analüüsimise ja kasutamise meetodite kohta (sealjuures töötuse kohta) on huvi tundmas ka Euroopa Komisjon, pakkudes Horison 2020 programmi raames rahastust vastavasisuliste meetodite loomiseks ning uuringute teostamiseks (Euroopa Komisjon 2016).

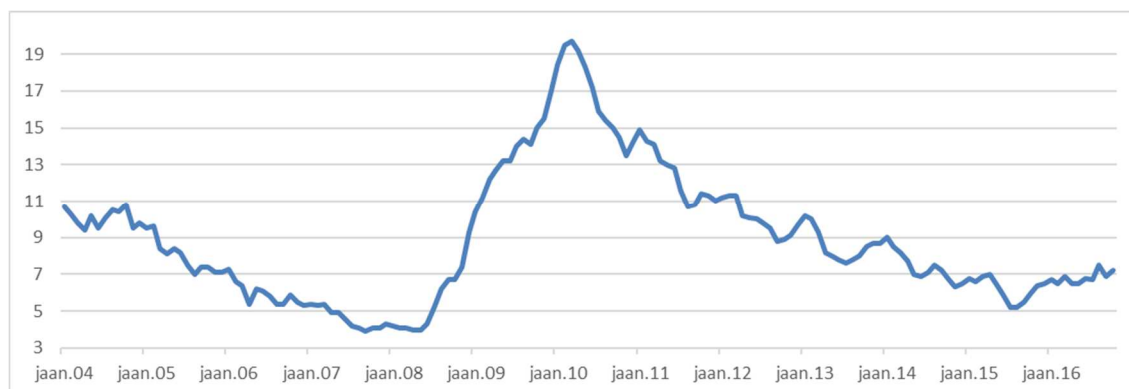
Paljude varasemate tööde fookus on olnud mõista, kas Google otsingute andmetel on potentsiaali prognoosimise parendamiseks. Käesolev töö püüab, lisaks eelnevale, välja pakkuda konkreetse tegevusjuhise, kuidas kaasata Google otsingute andmeid prognoosimudelitesse nii, et see vähendaks prognoosimise viga. Formuleeritava algoritmi abil on näiteks võimalik üles ehitada tarkvara rakendus, mis reaalajas tegeleks majandusnäitajate prognoosimisega. Käesolev töö on inspireeritud suurel määral Hyunyoung Choi, Hal Varian (2009a, 2009b ja 2011) ja Joonas Tuhkuri töödest (2014 ja 2016). Choi, Varian, Tuhkuri ja ka teiste autoritega võrreldes arendatakse prognoosimudeleid detailsemaks/tihedamaks ja teostatakse prognoosimine pikema tuleviku ajaperioodi (horisondi) kohta. Prognoosimudelite täiustused võimaldavad, võrreldes teiste autoritega (Tuhkuri 2014, 2016; Chadwick, Sengül 2012), prognoosida stabiilsemaid tulemusi, sealhulgas kaugema horisondi taha. Samuti stabiliseeritakse, erinevalt teistest töödest, Google Trends liialt hajuvaid päringutulemusi korduvate, erinevatel päevadel teostatud vaatluste läbiviimise abil.

2. ANDMED JA METOODIKA

2.1. Uuritavate näitajate statistilised andmed

Käesolevas töös vaadeldakse järgmisi Eesti majandusnäitajaid: töötuse määr, majutatute arv, jaemüük posti või Interneti teel ning kinnisvara ostu-müügitehingute arv. Näitajad on valitud sellepärast, et need on seotud majanduse üldise tervisliku olukorraga ja nende kohta on tehtud piisavalt otsinguid Google otsingumootoris. Järgnevalt antakse lühiülevaade näitajate definitsioonidest ja kirjeldatakse nende dünaamikat. Põhjalikum ülevaade näitajate statistiliste metoodikate kohta on toodud lisas 1. Näitajate statistiliste andmete allikana on toetunud Eestis riikliku statistika kogumise ja avalikustamisega tegeleva organisatsiooni, Statistikaameti, poolt kogutud andmetele. Käesolevas peatükis on ära toodud uuritavate näitajate jaanuar 2017 lõpuks teada olnud andmed.

Mõiste „**töötuse määr**“ all mõeldakse käesolevas töös töötute osatähtsust majanduslikult aktiivsest rahvastikust ehk tööjõust Eestis (ühes) kuus (vt ka lisa 1) protsentides (joonis 2). Eesti statistikaamet kuulisi töötuse määra andmeid ise ei avalda, küll aga edastab Statistikaamet need Eurostatile. Kuna antud töö raames on huvipakkuvad just töötuse määra igakuised andmed, siis viimased on leitavad Eurostati *une_rt_m* andmebaasist.



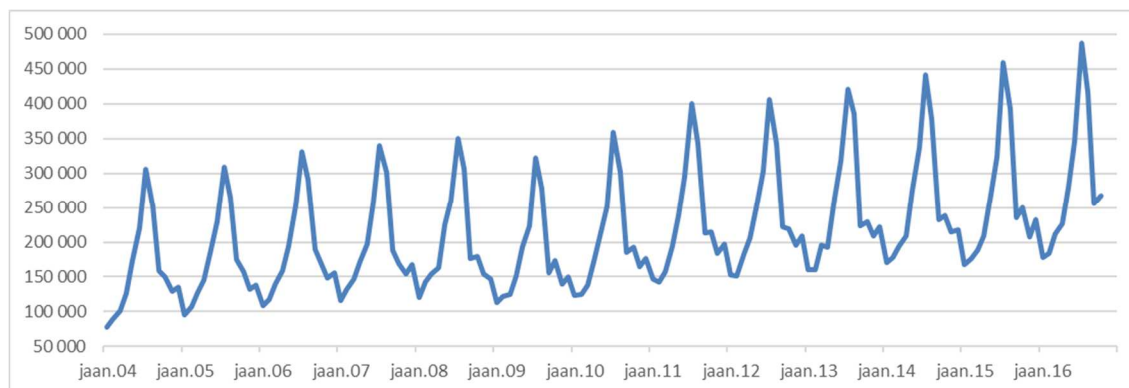
Joonis 2. Tegelik töötuse määr, %
(Allikas: Eurostat *une_rt_m* 2017)

Eurostati andmebaasi päring ja parameetrid:

Unemployment by sex and age: monthly average (une_rt_m)

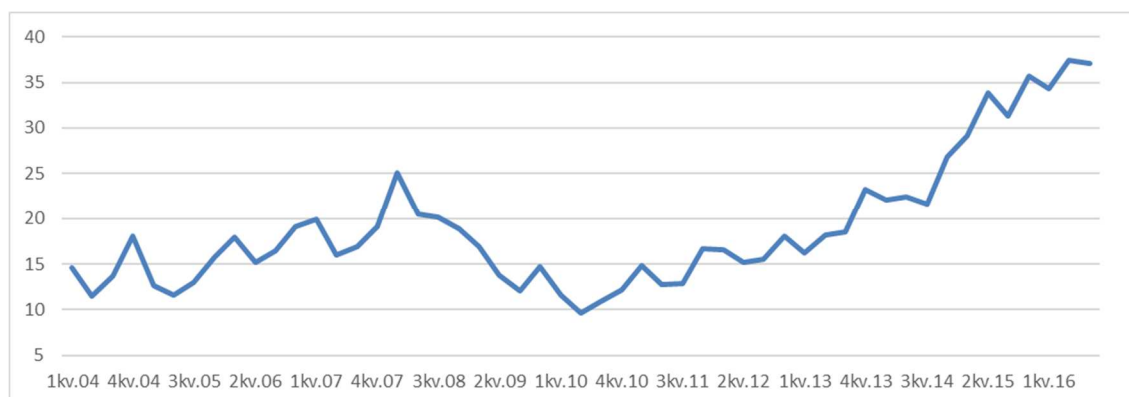
(S_ADJ: Unadjusted data; AGE: Total; UNIT: Percentage of active population; SEX: Total)

Mõiste „majutatute arv“ all mõeldakse käesolevas töös Eestis tegutsevates majutusettevõtetes majutatute koguarvu (ühes) kuus (vt ka lisa 1). Andmed on leitavad Statistikaameti TU121 andmebaasist.

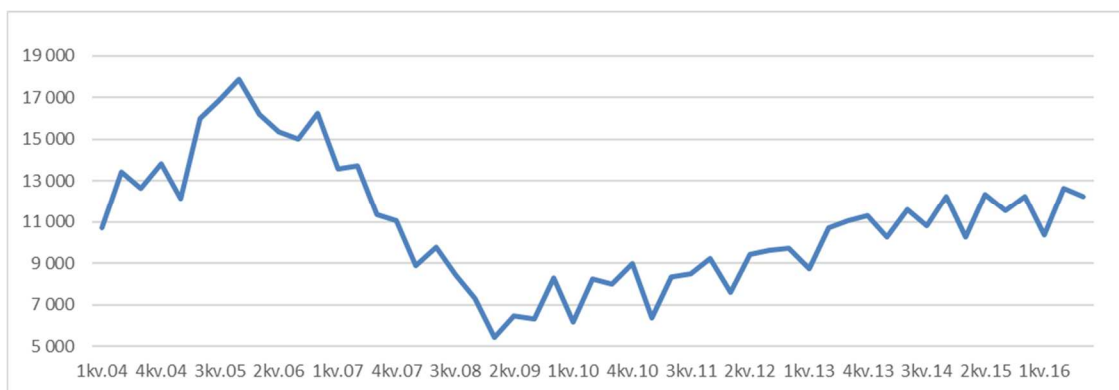


Joonis 3. Tegelik majutatute arv, tk
(Allikas: Statistikaamet TU121 2017)

Mõiste „jaemüük posti või Interneti teel“ all mõeldakse käesolevas töös Eestis tegutsevate ettevõtete EMTAK 2008 G47 (Jaekaubandus, v.a mootorsõidukid ja mootorrattad) G4791 (jaemüük posti või Interneti teel) koodi alusel tegevusalarühma jaekaubandusettevõtete müügitulu (ühes) kvartalis (vt ka lisa 1). Andmed on leitavad Statistikaameti KM0101 andmebaasist.



Mõiste „kinnisvara ostu-müügittehingute arvu“ all mõeldakse käesolevas töös kinnistusraamatusse kantud kinnisvara ostu-müügittehingute arvu (ühes) kvartalis (vt ka lisa 1). Andmed on leitavad Statistikaameti KV015 andmebaasist.



Joonis 5. Tegelik kinnisvara ostu-müügittehingute arv, tk
(Allikas: Statistikaamet KV015 2017)

Jaemüük posti või Interneti teel ning kinnisvara ostu-müügittehingute arvu näitajaid vaadeldakse kvartali, mitte kuu detailsusega sellepärast, et Statistikaamet ei avalda nende näitajate kohta kuu detailsusega andmeid.

2.2. Uuritavate näitajate Interneti-otsingute andmed ja Google indeksi koostamise meetoodika

Google Inc. Interneti otsingumootori (<https://www.google.ee/>) abil on võimalik otsida Internetist informatsiooni kasutades otsingutes erinevaid märksõnu ehk sõnu, millega seoses infot vajatakse. Populaarsemate kasutatud märksõnade kohta peab Google statistikat Google Trends nimelises rakenduses <https://www.google.com/trends/?hl=et>. Google Trends rakendusest on võimalik teha väljavõtteid: kui palju, millisest geograafilisest piirkonnast ja millises ajaperioodis (näiteks aastatel 2010-2016) on mõnda konkreetset märksõna Google otsingumootoris kasutatud. Statistika kuvatakse suhtarvuna, kus kõige suurem otsingute arv mingil ajahetkel (näiteks nädalas või kuus) võrdsustatakse 100-ga ja teistel ajahetkedel toimunud otsingute arv kuvatakse 100 punktises skaalal suhtarvuna. Antud meetoodika puuduseks on asjaolu, et kui otsingute märksõna maksimaalsed väärtused ajas muutuvad (aeg liigub edasi), siis ka statistika on ajas muutuv. Lisaks sõltuvad andmete arväärtused ajaperioodist, mille kohta huvi tuntakse, ajaperioodi granulaarsusest ehk detailsusest ning statistika väljavõtte

kuupäevast. Samuti on kahe samaväärse päringu tulemused erinevad olenevalt päringu võtmise ajast (sh päevast). Negatiivne aspekt on ka see, et me täpselt ei tea, kas ja kuidas arvestatakse statistika kuvamisel puhtalt Interneti-kasutajate arvu kasvamisest tulenevat päringute kasvamise trendi. Väljavõtteid on võimalik eksportida Google Trends rakendusest tekstidokumendi kujul (.csv formaat). Ehki autorile teadaolevalt ei paku Google ise Google Trends ametliku rakendusliidest (*Application Programming Interface API*), ei ole rakendusliidese loomine kasutades veebipäringuid ülemäära keeruline. Vastavaid vabavaralisi Google Trends rakendusliideseid (API) on võimalik ka Internetist all laadida (npm Inc.).

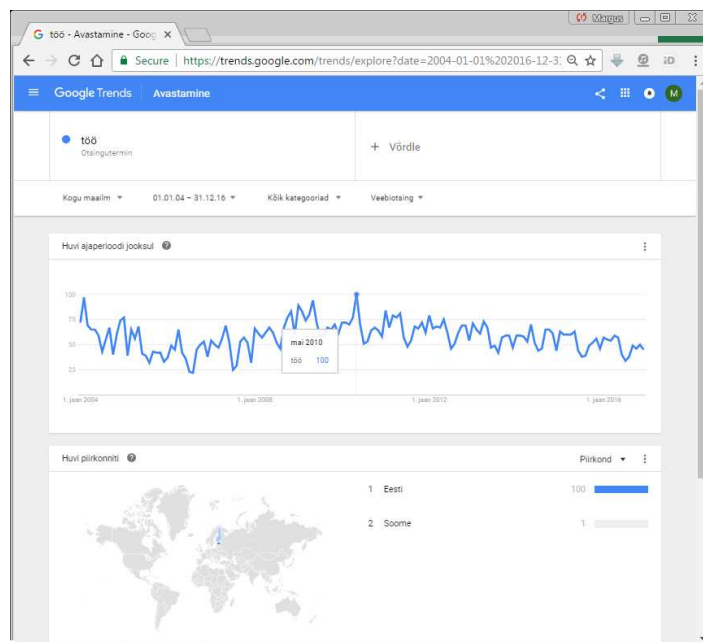
Väljavõtete andmete alusel moodustatakse käesoleva töö raames niinimetatud Google indeks (GI), mida kasutatakse prognoosimudelite ülesehitamiseks ja hiljem prognoosimiseks. Google indeks võib olla moodustatud ainult ühe märksõna või siis märksõnade komplekti baasil. Google indeks moodustatakse iga näitaja (töötuse määr, kinnisvara tehingute arv, jaemüük posti ja Interneti teel, majutatute arv) jaoks eraldi. Märksõnad ise ja nende arv on valitud selliselt, et moodustatud Google indeksi aegrea graafik oleks võimalikult sarnane konkreetse näitaja tegeliku aegrea graafikuga. Selleks kasutatakse tagasisidestatud loogika alusel tegeliku aegrea ja moodustatud Google indeksi aegrea graafilist analüüsi. Ehk, alustatakse põhimärksõnast ja kombineeritakse toetavaid märksõnu niikaua, kuni tegeliku aegrea ja Google indeksi aegrea graafikud oleksid võimalikult sarnased. Sarnasuse hindamisel kasutatakse võrdlemismeetodina keskmise absoluutse suhtelise vea meetodit – MAPE (*mean absolute percentage error*). Lisaks on võimalik hinnata graafikutes sisalduvate trendide sarnasusi visuaalselt. Märksõnade ja nende arvu valikul ehk aegride graafilise analüüsi teostamisel kasutatakse kõiki teadaolevaid tegelike ja Google märksõnade ajahetkede väärtuseid. Võrreldavuse huvides kasutatakse mineviku (*ex post*) ja tuleviku (*ex ante*) prognoosimisel (vt peatükk 2.3 ja 3.2) samu märksõnu ja nende arvu. Kuna Google Trend pakub märksõnade otsingute andmeid alates aastast 2004, siis ajaperioodide omavahelise võrreldavuse eesmärgil viiakse kogu analüüs ja arvutused läbi alates aastast 2004.

Selleks, et elimineerida märksõnade andmete ebastabiilsust, on võetud Google Trends abil märksõnade andmete väljavõtteid iga päev ühe kuu jooksul (jaanuar 2017). Olenevalt sellest, kas tegemist on mineviku või tuleviku prognoosimisega (vt peatükk 2.3) on

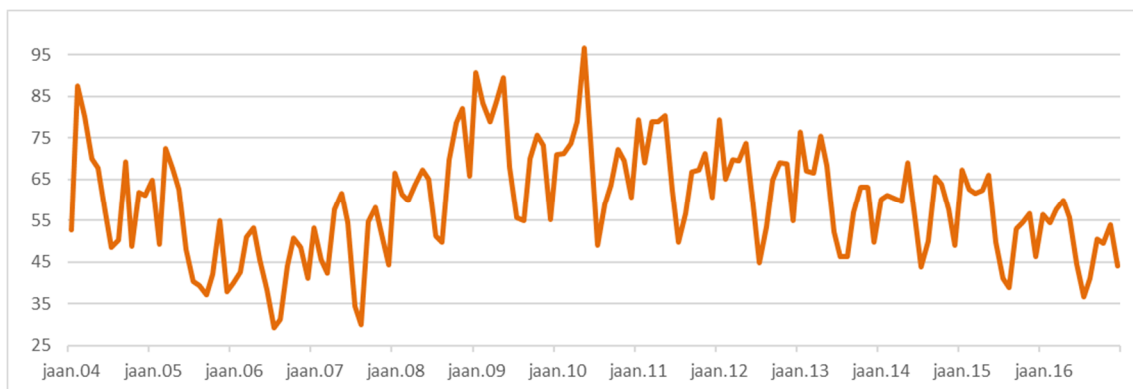
väljavõtete ajaperioodid erinevad. Iga päev võetud vaatlustulemused on salvestatud ja ajatempli eesmärgil digitaalselt allkirjastatud ning on salvestatud digitaalselt taas esitataval kujul. Google indeksi moodustamisel on võimalus arvestada Interneti kasutajate arvu kasvamisest tulenevat päringute kasvamise trendi. Lisaks on veel võimalus Google indeksi moodustamisel erinevaid märksõnu võtta arvesse erinevate kaaludega. Autorile teadaolevalt ei ole ükski varasem töö viimast kahte nimetatud võimalust Google indeksi moodustamisel kasutanud ja ka käesolevas töös ei ole neid arvesse võetud.

Igal päeval, 31 päeva jooksul, on salvestatud kaks komplekti vaatlustulemusi 25 märksõna kohta. Iga vaatlustulemus omab kas 156 (tuleviku prognoosimise puhul) või 141 (mineviku prognoosimise puhul) andmeühikut (konkreetne arv vahemikus 0-100) andmeid. Kokku on kogutud 1550 vaatlustulemusega 220875 andmeühikut Google Trends algandmeid.

Töötuse määra Google indeksi moodustamisel kasutatakse ainult ühte põhimärksõna „töö“ arvestades kogu maailmast teostatud päringuid. Töötuse määra Google indeks moodustatakse kuude kaupa. Konkreetse kuu väärtus saadakse, võttes kõikide päevade vaatlustulemuste väärtustest aritmeetiline keskmine.

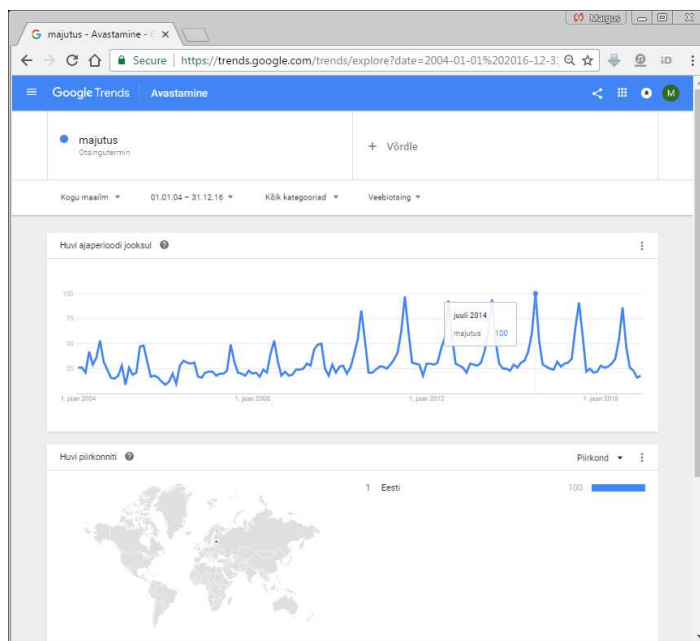


Joonis 6. Google Trends märksõna „töö“ 01.01.2014-31.12.2016 väljavõte (Allikas: Google Trends 2017)

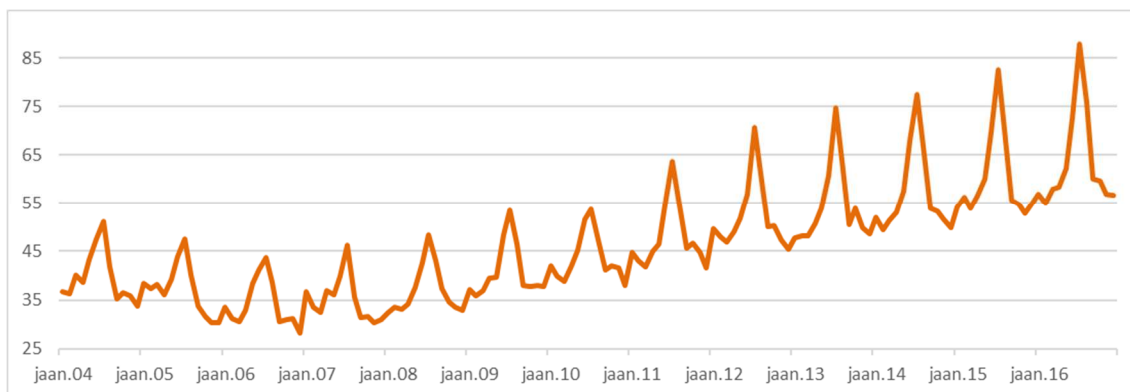


Joonis 7. Töötuse määra Google indeks (GI), suhtarv

Majutatute arvu Google indeksi moodustamist alustatakse põhimärksõnast „majutus“ ja lisamärksõnade kombineerimise tulemusena kaasatakse märksõnad: „hiiumaa“, „otepää“, „narva“, „нарва“, „päru“, „пяру“, „saaremaa“, „сааремаа“, „tartu“, „тарту“, „tallinn“, „таллинн“ ja „таллин“ arvestades kogu maailmast teostatud päringuid. Majutatute arvu Google indeks moodustatakse kuude kaupa. Konkreetse kuu väärtus saadakse, võttes kõikide päevade ja kõikide märksõnade vaatlustulemuste väärtustest aritmeetiline keskmine.

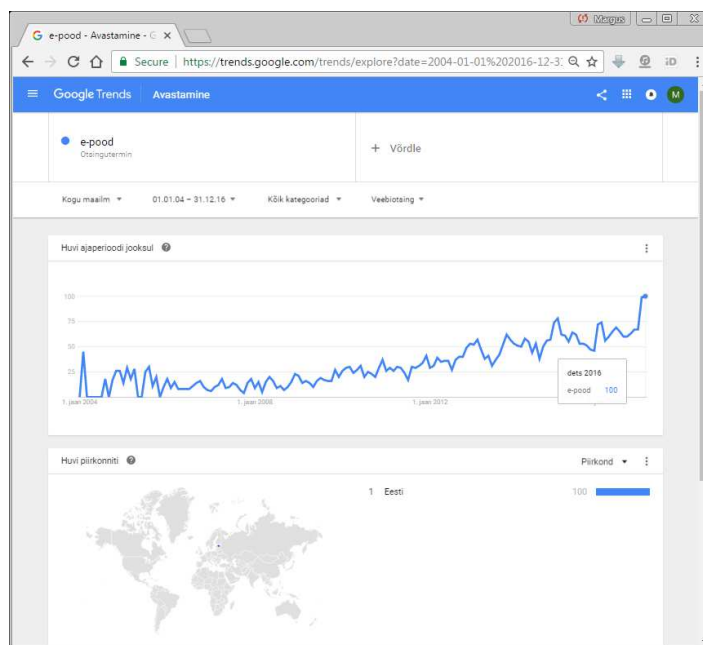


Joonis 8. Google Trends märksõna „majutus“ 01.01.2014-31.12.2016 väljavõte (Allikas: Google Trends)

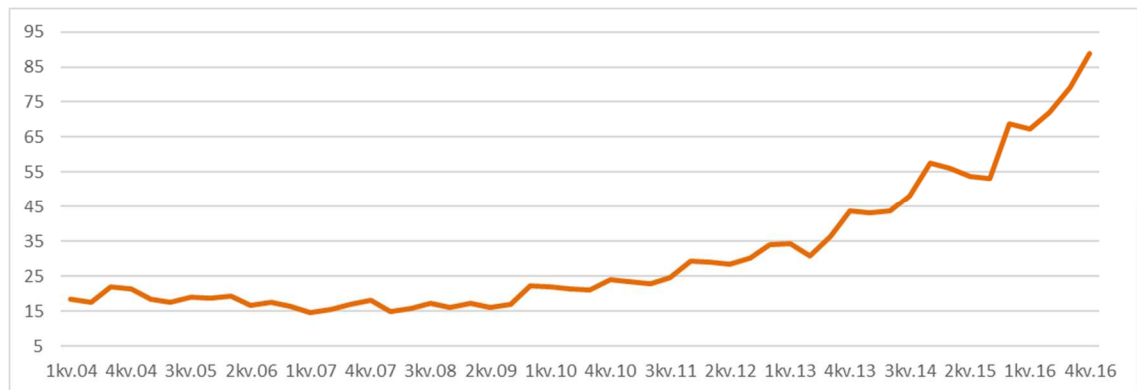


Joonis 9. Majutatute arvu Google Indeks (GI), suhtarv

Jaemüük posti või Interneti teel Google indeksi moodustamist alustatakse põhimärksõnast „epood“ ja lisamärksõnade kombineerimise tulemusena kaasatakse märksõnad: „e pood“, „e-pood“, „hansapost“ (mida on otsitud kogu maailmast) ja „ebay“, „amazon“, „aliexpress“ (mida on otsitud ainult Eestist). Jaemüük posti või Interneti teel Google indeks moodustatakse kvartalite kaupa. Konkreetse kvartali väärtus saadakse, võttes kõikide päevade ja kõikide märksõnade vastava kolme kuu vaatlustulemuste väärtustest aritmeetiline keskmine.

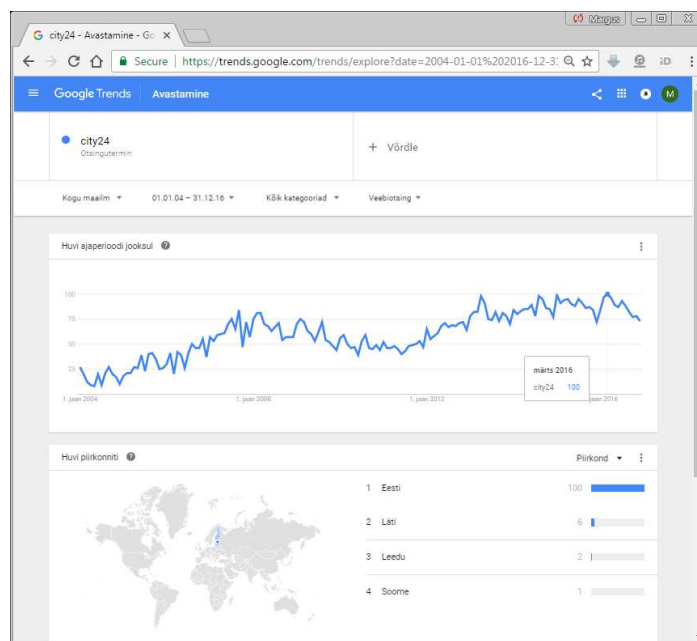


Joonis 10. Google Trends, märksõna „e-pood“ 01.01.2014-31.12.2016 väljavõte (Allikas: Google Trends)

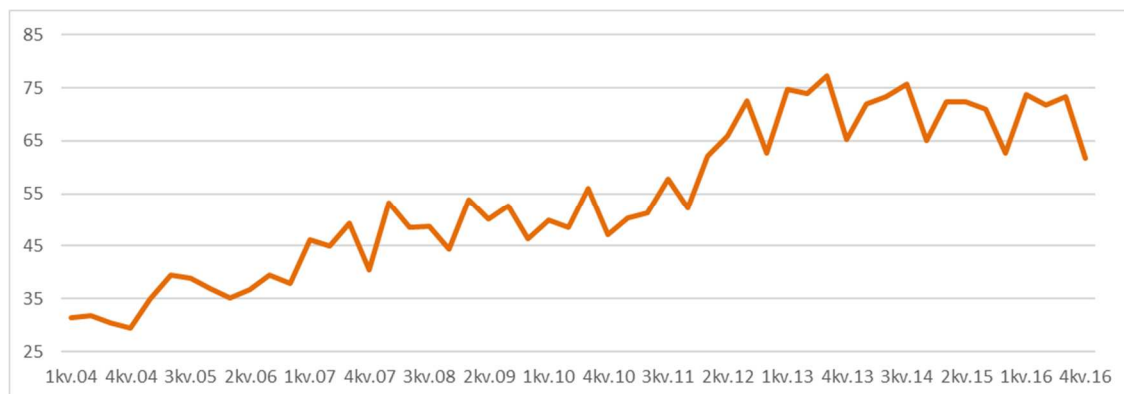


Joonis 11. Jaemüügi posti või Interneti teel Google indeks (GI), suhtarv

Kinnisvara ostu-müügitehingute arvu Google indeksi moodustamist alustatakse põhimärksõnast „kinnisvara“ ja lisamärksõnade kombineerimise tulemusena kaasatakse märksõnad: „city24“ ja „kv.ee“. Kinnisvara ostu-müügitehingute arvu Google indeks moodustatakse kvartalite kaupa. Konkreetse kvartali väärtus saadakse, võttes kõikide päevade ja kõikide märksõnade vastava kolme kuu vaatlustulemuste väärtustest aritmeetiline keskmine.



Joonis 12. Google Trends, märksõna „city24“ 01.01.2014-31.12.2016 väljavõte (Allikas: Google Trends)

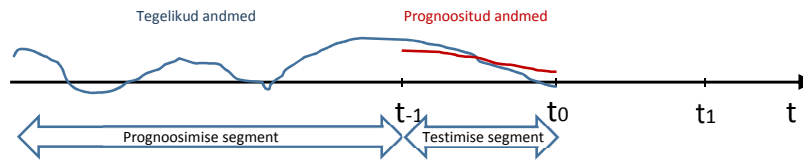


Joonis 13. Kinnisvara ostu-müügitehingute arvu Google indeks (GI), suhtarv

2.3. Prognoosimise meetodid

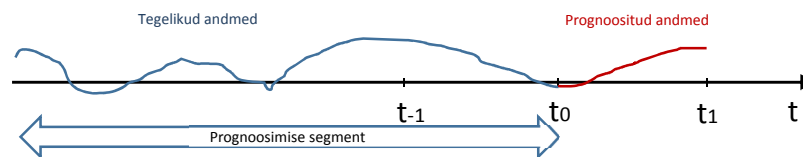
Erinevad prognoosimise ülesande püstitused vajavad erinevaid lähenemisi prognoosimise mudelite ülesehituseks. Antud töö eesmärki silmas pidades on vaatluse all prognoosimine, mille läbiviimise tulemusel on võimalik anda praktilist laadi informatsiooni andmete kohta, mida eelduslikult me veel ei tea. Alljärgnevas osas antakse lühiülevaade võimalikest prognoosimise lähenemistest. Prognoosimise jaotust vaadeldakse mingi konkreetse prognoosimise omaduse vaates. Alljärgnevate peatükkide koostamisel on kasutatud Yaffee (2010) ja Kangro (2010) töid.

Mineviku prognoosimist (*ex post*) viiakse läbi selleks, et kaaluda prognoosimudeli adekvaatsust juba minevikust teadaolevate tegelike andmete suhtes. Selleks jagatakse info kaheks segmendiks. Esimesse segmenti kuuluvad tegelikud andmed, mille alusel prognoosimudel üles ehitatakse – prognoosimise segment ehk ajalooline segment (*estimation* ehk *historical* segment). Teise segmenti kuuluvad samuti tegelikud andmed, kuid nende andmete suhtes teostatakse loodud prognoosimudeli headuse proov – hindamise ehk testimise segment (*validation* ehk *test* segment). Teisisõnu prognoositakse prognoosimise segmendi andmete alusel testimise segmendi andmeid ja võrreldakse viimaseid tegelike andmetega. Võrdlusest tuleneb prognoosimise viga. Prognoosimise vea abil on võimalik hinnata loodud prognoosimudeli adekvaatsust.



Joonis 14. Mineviku (*ex post*) prognoosimine
(Allikas: Yaffee 2010)

Tuleviku prognoosimist (*ex ante*) viiakse läbi juhtudel, kus meil on praktiline vajadus prognoosida andmeid, mida me veel tegelikult ka ei tea. Prognoosimudelisse kaasatakse kõik alusandmed, mis on selleks hetkeks juba teada. Kui meil ei ole mingit kokkuleppelist malli või standardit, millega prognoosimise tulemusi võrrelda, siis ei ole võimalik leida ka prognoosimise viga ega anda hinnangut prognoosimudeli adekvaatsuse kohta. Sellest probleemist ülesaamiseks võib koostada antud näitele kohandatud kokkuleppelise „naturaalse“ prognoosi, mille suhtes on seejärel võimalik prognoose võrrelda. Sellisel juhul on võimalik välja arvutada prognoosimise viga ja hinnata kasutatud prognoosimudeli adekvaatsust.



Joonis 15. Tuleviku (*ex ante*) prognoosimine
(Allikas: Yaffee 2010)

Käesolevas töös viiakse läbi mineviku prognoosimine (*ex post*), arvutatakse *ex post* prognoosimise viga võrreldes tegelike andmetega ja hinnatakse kasutatud prognoosimudelite ja seeläbi prognoosimise ülesehituse loogika adekvaatsust. Seejärel koostatakse sama prognoosimise ülesehituse loogika alusel tuleviku (*ex ante*) prognoosid. Kuna näidetele kohandatud „naturaalseid“ prognoose ei koostata, siis *ex ante* prognoosidele prognoosimise viga ka välja ei arvutata. Eeldatakse, et kui mineviku prognoosimise (*ex post*) ülesehitamise loogika on adekvaatne, siis on samaväärse loogika alusel ülesehitatud tuleviku prognoosimine (*ex ante*) ka piisavalt adekvaatne, et välja pakkuda tuleviku prognoosid.

Ennem, kui selgitada staatilise (*static*) ja dünaamilise (*dynamic*) prognoosimise erinevust, tuleb vaadelda üks-samm-ette (*one-step-ahead*) versus mitu-sammu-ette (*multi-stap-*

ahead) prognoosimise vahet. Kui-üks-samm ette prognoosimisel prognoositakse ainult vahetult järgnevat ajahetke, siis mitu-sammu-ette prognoosimisel prognoositakse mitut ajahetke (*horizon* ehk horisont) alates baasajahetkest. Staatile ja dünaamiline prognoosimine on sarnane, kui tegemist on üks-samm-ette prognoosimisega – kasutatakse prognoosimiseks ju kõike tegelikke andmeid. Vahe tuleb sisse siis, kui tuleb prognoosida mitu-sammu-ette. Staatile prognoosimine võtab arvesse ainult tegelikke andmeid, mille põhjal teostatakse mitme ajahetke prognoos. Dünaamiline prognoosimine võtab arvesse tegelikke andmeid ja eelnevatel sammudel juba prognoositud andmeid, mille põhjal teostatakse järgmise ajahetke prognoos. Käesolevas töös kasutatakse staatilist mitu-sammu-ette prognoosimist.

Ühemõõtmeline (*unidimensional*) prognoosimine teostatakse juhul, kui kasutatakse ainult vaadeldava aegrea minevikuväärtuseid. Mitmemõõtmelist (*multidimensional*) prognoosimist on võimalik teha siis, kui lisaks vaadeldava aegrea minevikuväärtustele on võimalik kaasata ka väliseid andmeid, näiteks mõne muu aegrea minevikuväärtuseid. Käesolevas töös nimetatakse varasemate töödega seose säilimiseks ühemõõtmeline prognoosimudel baasmudeliteks ja mitmemõõtmeline prognoosimudel laiendatud mudeliteks.

Antud töö toetub suurel määral Hyunyoung Choi, Hal Varian (2009a) ja Joonas Tuhkuri prognoosimudelitele (2016). Choi, Variani (2009a) mudelid avalduvad valem 1. (baasmudel) ja valem 2. (laiendatud mudel) kujul. Autentsuse huvides on säilitatud autorite valemite kirjapilt.

$$(1) \text{ Mudel 0: } \log(y_t) \sim \log(y_{t-1}) + \log(y_{t-12}) + e_t,$$

$$(2) \text{ Mudel 1: } \log(y_t) \sim \log(y_{t-1}) + \log(y_{t-12}) + x_t^{(1)} + e_t$$

kus y_t – tegeliku aegrea väärtus ajahetkel t ,
 y_{t-1} – tegeliku aegrea väärtus ajahetkel $t - 1$,
 y_{t-12} – tegeliku aegrea väärtus ajahetkel $t - 12$,
 $x_t^{(1)}$ – (esimese) välise aegrea (Google indeks) väärtus ajahetkel t ,
 e_t – viga.

Konkreetsete näitajate arvutustel leitakse prognoosimudeli parameetrid. Autorid viitavad, et selliseid mudeleid nimetatakse kirjanduses sesoonseteks autoregresiivseteks mudeliteks või sesoonseteks AR mudeliteks (*seasonal autoregressive model* või *seasonal AR model*). Nagu näha on konkreetsed viiteaegadega mudelid mittelineaarsel ehk kümnendlogaritmilisel kujul.

Tuhkuri (2016) toetub oma prognoosimudelites Choi, Variani mudelitele. Mudelid avalduvad valem 3. (baasmudel) ja valemi 4. (laiendatud mudelid) komplekti kujul. Autentsuse huvides on säilitatud autori valemite kirja pilt.

(3) Mudel (0.0): $\log(y_t) = \beta_0 + \beta_1 \log(y_{t-1}) + \beta_2 \log(y_{t-12}) + e_t$

Mudel (1.0): $\log(y_t) = \beta_{00} + \beta_{10} \log(y_{t-1}) + \beta_{20} \log(y_{t-12}) + \beta_{30} x_t + e_t$

Mudel (1.1): $\log(y_t) = \beta_{01} + \beta_{11} \log(y_{t-1}) + \beta_{21} \log(y_{t-12}) + \beta_{31} x_{t-1} + e_t$

Mudel (1.1): $\log(y_t) = \beta_{02} + \beta_{12} \log(y_{t-1}) + \beta_{22} \log(y_{t-12}) + \beta_{32} x_{t-2} + e_t$

(4) Mudel (1.3): $\log(y_t) = \beta_{03} + \beta_{13} \log(y_{t-1}) + \beta_{23} \log(y_{t-12}) + \beta_{33} x_{t-3} + e_t$

Mudel (1.4): $\log(y_t) = \beta_{04} + \beta_{14} \log(y_{t-1}) + \beta_{24} \log(y_{t-12}) + \beta_{34} x_{t-4} + e_t$

Mudel (1.5): $\log(y_t) = \beta_{05} + \beta_{15} \log(y_{t-1}) + \beta_{25} \log(y_{t-12}) + \beta_{35} x_{t-5} + e_t$

Mudel (1.6): $\log(y_t) = \beta_{06} + \beta_{16} \log(y_{t-1}) + \beta_{26} \log(y_{t-12}) + \beta_{36} x_{t-6} + e_t$

kus y_t – tegeliku aegrea väärtus ajahetkel t ,
 y_{t-1} – tegeliku aegrea väärtus ajahetkel $t - 1$,
 y_{t-12} – tegeliku aegrea väärtus ajahetkel $t - 12$,
 x_t – välise aegrea (Google indeks) väärtus ajahetkel t jne.,
 β_{0h} kuni β_{3h} – mudeli parameetrid,
 e_t – viga.

Konkreetsete näitajate arvutustel leitakse prognoosimudeli parameetrid β . Tuhkuri toob sisse üks-samm-ette ja mitu-sammu-ette prognoosimisel horisondi ($h=6$) mõiste teostades lisaks üks-samm-ette prognoosimisele ka dünaamilise h -sammu-ette prognoosimise. Joonas Tuhkuri tööde prognoosid on toodud lisas 3. Võrreldes Choi, Varian (2009a) ja Tuhkuri (2016) mudelitega on käesolevas töös prognoosimudeleid tihendatud ja horisonti

pikendatud. Samuti on avaldatud välise aegrea mõju mittelineaarsel kujul (vt peatükk 3.2). Prognoosimudeli tihendamise all mõeldakse seda, et võrreldes Choi, Varian (2009a) ja Tuhkuri (2016) töödes kasutatud mudelitega on mudelite komponentide arvu esimese ($t-1$) ja viimase ($t-12$) komponendi vahel suurendatud nii, et arvesse on võetud iga kuu või kvartali andmed esimesest kuni viimase komponendini. Horisondi pikendamise all mõeldakse seda, et võrreldes Tuhkuri (2016) töödega, kus prognoosimine teostati seitsme ajahetke kohta (üks oleviku + kuus tuleviku ajahetke), siis käesolevas töös teostatakse prognoosimine kuni viieteistkümne ajahetke kohta (kaks mineviku + üks oleviku + kaksteist tuleviku ajahetke) (vt ka lisa2 ja lisa 3).

3. PROGNOOSIMUDELITE HINDAMINE JA TULEMUSED

3.1. Aegridade graafiline analüüs

Selleks, et tajuda tegelike aegridade ja Interneti-otsingute arvu baasil moodustatud Google indeksi aegridade väärtusi ning nende omavahelist suhet, teostatakse aegridade graafilised võrdlused. Graafilist analüüsi kasutatakse märksõnade ja nende arvu valikul. Sarnasust hinnatakse kõigepealt visuaalselt ja seejärel kasutatakse võrdlemismeetodina keskmise absoluutse suhtelise vea meetodit – MAPE (*mean absolute percentage error*). Alljärgnevaid Google indeksi matemaatilisi teisendusi kasutatakse ainult aegridade graafiliseks analüüsiks. Prognoosimisel kasutatakse originaalset Google indeksit.

$$(5) \quad MAPE = \frac{1}{t_2 - t_1} \sum_{t=t_1}^{t_2} \left| \frac{(x_{s\,t} - y_t)}{y_t} \times 100 \right|$$

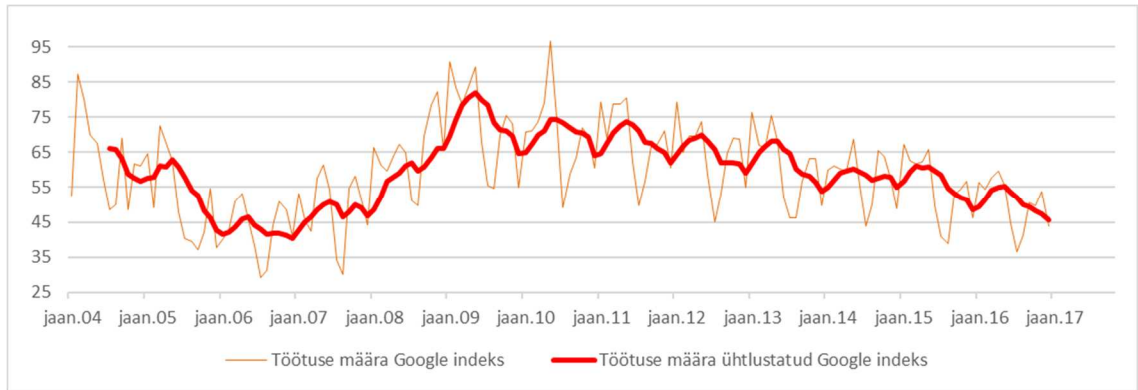
kus $x_{s\,t}$ – skaleeritud Google indeksi aegrea arvuline väärtus ajahetkel t , töötuse
määra näite korral asendatakse valemis $x_{s\,t}$ väärtused $x_{n\,t}$ väärtustega,
 y_t – tegeliku aegrea väärtus ajahetkel t ,
 t_1 – esimene ühine võrreldav ajahetk,
 t_2 – viimane ühine võrreldav ajahetk.

Tegeliku töötuse määra ja töötuse määra Google indeksi aegridade graafiliseks analüüsiks viiakse läbi kolm matemaatilist teisendust.

1. Töötuse määra Google indeksi aegrea silumine ehk ühtlustamine nn. libiseva keskmise meetodil. Iga ühtlustatud aegrea väärtus saadakse sellele eelneva kuue kuu pluss jooksva kuu keskmise võtmise tulemusena. Teisenduse eesmärk on kaotada Google indeksi aegrea liigne hajuvus (joonis 16).

$$(6) \quad x_{\bar{u}\,t} = \frac{1}{7} \sum_{i=0}^6 (x_{t-i})$$

kus $x_{\ddot{u}t}$ – ühtlustatud Google indeksi väärtus ajahetkel t ,
 x_{t-i} – Google indeksi väärtus ajahetkel t miinus i ,
 ehk antud näite juures käesolev kuu ja kuus eelnevat kuud,
 t – ajahetk, antud näitel on esimene ajahetk seismes kuu, ehk juuli 2004.

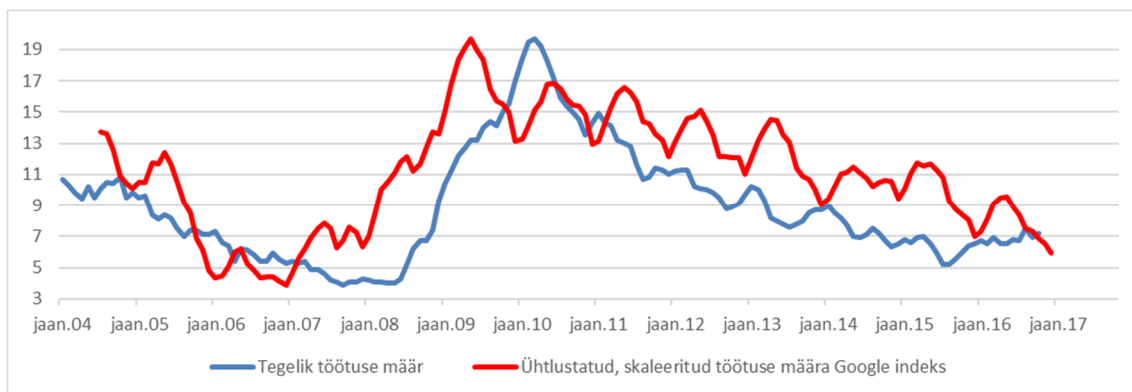


Joonis 16. Töötuse määr Google indeksi graafik koos ühtlustatud töötuse määr Google indeksi graafikuga, suhtarv

2. Ühtlustatud Google indeksi aegrea skaleerimine tegeliku töötuse määr aegrea suhtes. Ehk teisisõnu skaleeritakse Google indeksi graafik vertikaalselt nii, et ühtlustatud Google indeksi väärtused vahemikus 38,3 kuni 79,9 muutuksid töötuse määr väärtuse vahemikku 3,9 kuni 19,7. Teisenduse eesmärk on tuua töötuse määr ühtlustatud Google indeksi graafik tegeliku töötuse määr graafikuga samasse mõõtkavasse (joonis 17).

$$(7) \quad x_{st} = y_{tmin} + \frac{(x_{\ddot{u}t} - x_{\ddot{u}min}) \times (y_{tmax} - y_{tmin})}{(x_{\ddot{u}max} - x_{\ddot{u}min})}$$

kus x_{st} – Google indeksi skaleeritud väärtus ajahetkel t ,
 y_{tmin} – tegeliku töötuse määr minimaalne väärtus,
 $x_{\ddot{u}t}$ – ühtlustatud Google indeksi väärtus ajahetkel t ,
 $x_{\ddot{u}min}$ – ühtlustatud Google indeksi minimaalne väärtus,
 y_{tmax} – tegeliku töötuse määr maksimaalne väärtus,
 $x_{\ddot{u}max}$ – ühtlustatud Google indeksi maksimaalne väärtus.



Joonis 17. Töötuse määra ühtlustatud ja skaleeritud Google indeksi graafik koos tegeliku töötuse määra graafikuga, %

3. Ühtlustatud ja skaleeritud Google indeksi aegrea graafiku nihutamine, et tegeliku töötuse määra ja Google indeksi graafikute maksimumid satuksid samasse ajahetke. Teisenduse eesmärk on nihutada Google indeksi skaleeritud aegrea graafikut horisontaalselt niimitme ajaperioodi võtta, et

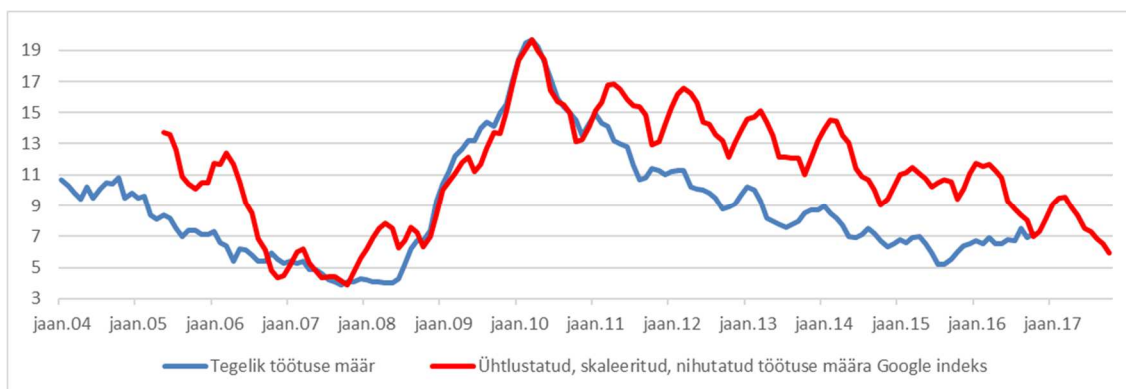
$$x_{\ddot{u} \max} = y_{t \max}$$

Antud näitaja puhul 10 kuud, ehk:

$$(8) \quad x_{n t} = x_{\ddot{u} t-10}$$

kus $x_{n t}$ – nihutatud Google indeksi väärtus ajahetkel t ,

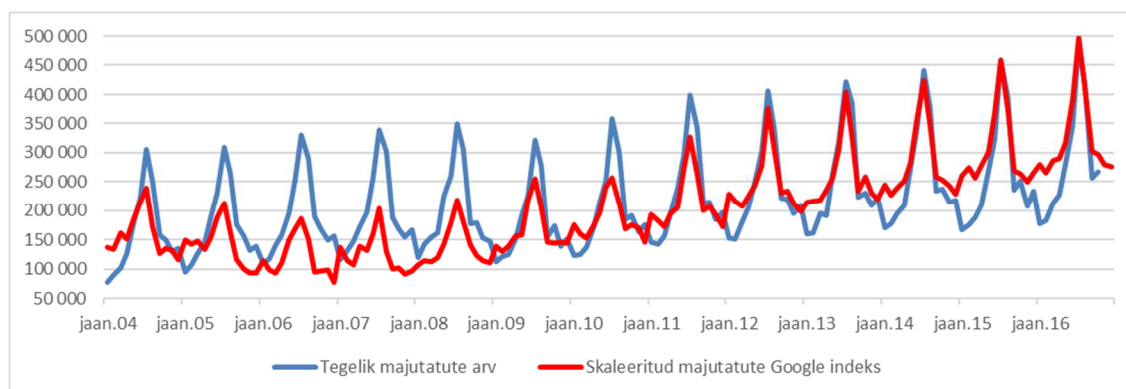
t – ajahetke väärtus, antud näitel on esimene ajahetk kuueteistkümnes kuu, ehk mai 2005.



Joonis 18. Töötuse määra ühtlustatud, skaleeritud ja nihutatud Google indeksi graafik koos tegeliku töötuse määra graafikuga, %

Joonisel 18 on võimalik märgata tegeliku töötuse määra ja töötuse määra Google indeksi graafikute sarnasusi, kus arvuliste väärtuste kasvu- ja kahanemistrendid satuvad samadesse ajaperioodidesse. See annab eelduse oletamaks, et töötuse määra Google indeksi aegreast võib abi olla töötuse määra prognoosimisel. Tegeliku töötuse määra väärtuste ja töötuse määra ühtlustatud, skaleeritud ja nihutatud Google indeksi väärtuste vaheline keskmine absoluutne suhteline viga (MAPE) on 38,3%.

Tegeliku majutatute arvu ja majutatute arvu Google indeksi aegridade graafiliseks võrdlemiseks viiakse läbi üks matemaatiline teisendus. Teostatakse majutatute arvu Google indeksi aegrea skaleerimine tegelikult majutatute arvu aegrea suhtes. Ehk teisistõnu skaleeritakse Google indeksi aegrea graafik vertikaalselt. Teisenduseks kasutatakse valemit 7 kohandades seda antud näitaja jaoks. Teisenduse eesmärk on tuua majutatute arvu Google indeksi graafik tegeliku majutatute arvu graafikuga samasse mõõtkavasse.

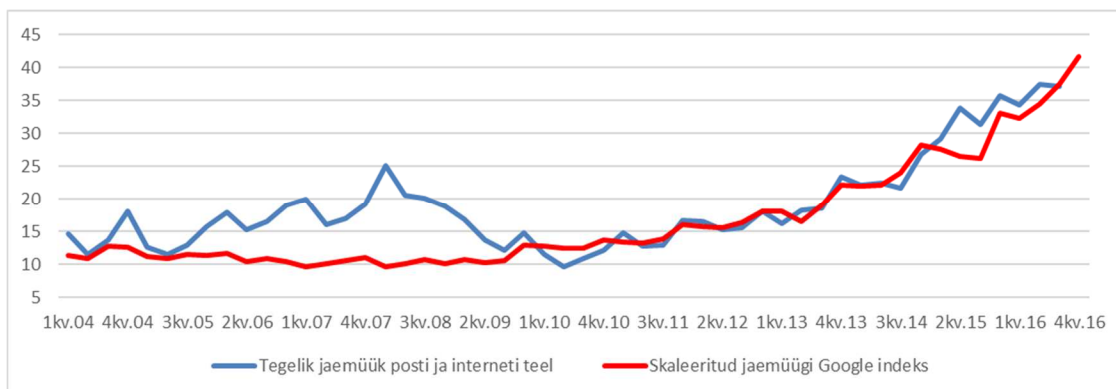


Joonis 19. Skaleeritud majutatute arvu Google indeksi graafik koos tegeliku majutatute arvu graafikuga, tk

Joonisel 19 on võimalik märgata tegelikult majutatute arvu ja skaleeritud Google indeksi graafikute sarnasusi. See annab eelduse oletamaks, et moodustatud majutatute arvu Google indeksi aegreast võib abi olla majutatute arvu prognoosimisel. Tegeliku majutatute arvu väärtuste ja skaleeritud majutatute arvu Google indeksi väärtuste vaheline keskmine absoluutne suhteline viga (MAPE) on 20,8%.

Tegeliku jaemüügi posti või Interneti teel ja jaemüügi posti või Interneti teel Google indeksi aegridade graafiliseks võrdlemiseks viiakse läbi analoogselt eelmise näitega Google indeksi skaleerimine kohandades valemit 7. Teisenduse eesmärk on tuua

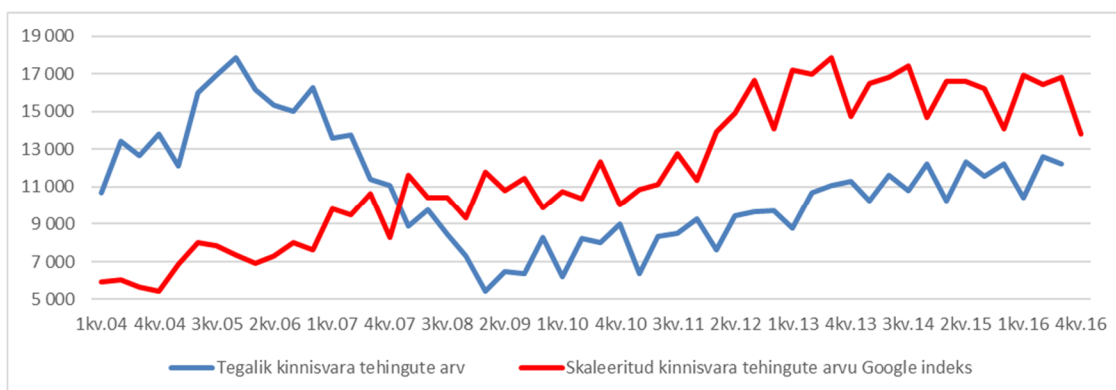
jaemüügi posti või Interneti teel Google indeksi graafik tegeliku jaemüügi posti või Interneti teel graafikuga samasse mõõtkavasse.



Joonis 20. Skaleeritud jaemüügi posti või Interneti teel Google indeksi graafik koos tegeliku jaemüügi posti või Interneti teel graafikuga, mln. eurot

Jooniselt 20 on võimalik märgata tegeliku jaemüügi posti või Interneti teel ja skaleeritud Google indeksi graafikute sarnasusi. See võimaldab oletada, et Google indeksi aegreast võib abi olla jaemüügi posti või Interneti teel prognoosimisel. Tegeliku jaemüügi väärtuste ja vastava skaleeritud Google indeksi väärtuste vaheline keskmine absoluutne suhteline viga (MAPE) on 18,5%.

Tegeliku kinnisvara ostu-müügitehingute arvu ja kinnisvara ostu-müügitehingute arvu Google indeksi aegridade graafiliseks võrdlemiseks kasutatakse samuti Google indeksi skaleerimist (valem 7). Teisenduse eesmärk on tuua kinnisvara ostu-müügitehingute arvu Google indeksi graafik tegeliku kinnisvara ostu-müügitehingute arvu graafikuga samasse mõõtkavasse.



Joonis 21. Skaleeritud kinnisvara ostu-müügitehingute arvu Google indeksi graafik koos tegeliku kinnisvara ostu-müügitehingute arvu graafikuga, tk

Jooniselt 21 on võimalik märgata tegeliku kinnisvara ostu-müügitehingute arvu ja skaleeritud kinnisvara ostu-müügitehingute arvu Google indeksi graafiku sarnasust alates aastast 2009. See võimaldab oletada, et Google indeksi aegreast võib abi olla kinnisvara ostu-müügitehingute arvu prognoosimisel. Tegeliku kinnisvara ostu-müügitehingute arvu väärtuste ja skaleeritud kinnisvara ostu-müügitehingute arvu Google indeksi väärtuste vaheline keskmine absoluutne suhteline viga (MAPE) on 47,2%.

3.2. Prognoosimudelid

Käesolevas töös kasutatakse prognoosimiseks tavalisi viiteaegadega mudeleid. Töö baasajahetk on jaanuar 2017. Selleks ajahetkeks on teada kvartali detailsusega näitajate puhul kolmanda kvartali 2016 tegelikud andmed ja kuu detailsusega näitajate puhul oktoobri 2016 tegelikud andmed. Google indeks on võimalik konstrueerida jaanuar 2004 kuni detsember 2016 kohta.

Töös peetakse tuleviku ajahetkede all silmas kõiki neid ajahetki, mille kohta ei ole veel teada tegelikke väärtuseid. Siia alla kuulub ka jooksev (olevik) ja lähimineviku ajahetk(ed).

Kvartali detailsusega näitajate puhul on eesmärk välja pakkuda 2016 aasta neljanda kvartali ja 2017 aasta kõikide kvartalite prognoosid (üks mineviku + üks oleviku + kolm tuleviku ehk kokku viis ajahetke). Google indeks konstrueeritakse esimene kvartal 2004 kuni neljas kvartal 2016 kohta.

Kuu detailsusega näitajate puhul on eesmärk välja pakkuda 2016 aasta kahe viimase kuu ja 2017 aasta kõikide kuude prognoosid (kaks mineviku + üks oleviku + kaksteist tuleviku ehk kokku viisteist ajahetke). Google indeks konstrueeritakse jaanuar 2004 kuni detsember 2016 kohta.

Enne tulevikuväärtuste prognoosimist tuleb aga veenduda, et prognoosimudelite ülesehituse loogika omab mõtet. Selleks läheme ajas tagasi ja koostame prognoosimudelid nii, et nende väljapakutud (prognoosimise segmendi põhiseid) prognoose on võimalik võrrelda minevikust teadaolevate tegelike andmete (test-segmendi) suhtes. Mineviku prognoosimudelite koostamisel kaasatakse ainult prognoosimise segmendi ehk vanemad tegelikud andmed kuni mineviku

prognoosimodeli baasajahetkeni. Testsegmenti andmed on prognoosimodeli baasajahetkest värskemad tegelikud andmeid.

Kvartali detailsusega näitajate puhul on vajadus ajas tagasi pöörduda neli ajahetke alates kolmandast kvartalist 2016 ehk neljandasse kvartalisse 2015 (kvartali detailsusega mineviku prognoosimodelite baasajahetk). Seega koostatakse mineviku prognoosid kolmas kvartal 2015 kuni kolmas kvartal 2016 kohta (kokku viis ajahetke). Google indeks konstrueeritakse esimene kvartal 2004 kuni kolmas kvartal 2015 kohta.

Kuu detailsusega näitajate puhul on vajadus ajas tagasi pöörduda viisteist ajahetke alates detsembrist 2017 ehk oktoobrisse 2015 (kuu detailsusega mineviku prognoosimodelite baasajahetk). Seega koostatakse mineviku prognoosid august 2015 kuni oktoober 2016 kohta (kokku viisteist ajahetke). Google indeks konstrueeritakse jaanuar 2004 kuni september 2016 kohta.

Mineviku prognoosimisel on võimalik võrrelda prognoosimodeli poolt väljapakutud väärtuseid tegelike mineviku väärtustega. On võimalik välja arvutada prognoosimodeli ja tegelike väärtuste omavaheline prognoosimise viga iga ajahetke kohta. Lõplik hinnang saadakse, võttes ajaperioodis kokku kõikidele ajahetkedele arvutatud vigade väärtustest keskmine. Antud töös kasutatakse prognoosimodeli poolt väljapakutud arväärtuste ja tegelike mineviku arväärtuste võrdlemismeetodina keskmise absoluutse suhtelise vea meetodit – MAPE (*mean absolute percentage error*).

$$(9) \quad MAPE = \frac{1}{t_2 - t_1} \sum_{t=t_1}^{t_2} \left| \frac{(y_{p\ t} - y_t)}{y_t} \times 100 \right|$$

kus $y_{p\ t}$ – prognoositud aegrea arvuline väärtus ajahetkel t ,

y_t – tegeliku aegrea väärtus ajahetkel t ,

t_1 – esimene prognoositud ajahetk,

t_2 – viimane prognoositud ajahetk.

Prognoosimudeleid on võimalik ülesse ehitada kahe põhimõttelise erinevusega:

1. Baasmudelid – prognoosimudel võtab arvesse ainult tegelikke eelmiste perioodide arvvaartuseid.
2. Laiendatud mudelid – prognoosimudel võtab arvesse tegelikke eelmiste perioodide arvvaartuseid ning lisaks (väliseid) Google indeksi eelmiste perioodide arvvaartuseid.

Mineviku prognoosimisel, kui laiendatud mudeli prognoosimise viga (võrreldes tegelike väärtustega) on väikesem kui baasmudelil, siis võib väita, et Google indeksi kaasamine prognoosimudelisse omab mõtet ehk parandab prognoosimise täpsust.

Käesolevas töös tuleviku prognoosimisel baasmudelite ja laiendatud mudelite prognoosimise viga välja ei arvutata, kuna antud näitele kohandatud malli ega kokkuleppelist „naturaalset“ prognoosi ei koostata.

Kvartali detailsusega baasmudelid avalduvad kujul:

$$(10) \quad M(0.h) : \log(y_{p\ t}) = \beta_{0h} + \sum_{i=1}^4 \beta_{ih} \log(y_{t-(h+(i-1))})$$

Kvartali detailsusega laiendatud mudelid avalduvad kujul:

$$(11) \quad M(1.h) : \log(y_{p\ t}) = \alpha_{0h} + \sum_{i=1}^4 \alpha_{ih} \log(y_{t-(h+(i-1))}) + \sum_{j=0}^4 \gamma_{jh} \log(x_{t-(h+(j-1))})$$

Kuu detailsusega baasmudelid avalduvad kujul:

$$(12) \quad M(0.h) : \log(y_{p\ t}) = \beta_{0h} + \sum_{i=1}^{12} \beta_{ih} \log(y_{t-(h+(i-1))})$$

Kuu detailsusega laiendatud mudelid avalduvad kujul:

$$(13) \quad M(1.h) : \log(y_{p\ t}) = \alpha_{0h} + \sum_{i=1}^{12} \alpha_{ih} \log(y_{t-(h+(i-1))}) + \sum_{j=-1}^{12} \gamma_{jh} \log(x_{t-(h+(j-1))})$$

kus $y_{p\ t}$ – prognoositud aegrea arvuline väärtus ajahetkel t ,
 y_t – tegeliku aegrea väärtus ajahetkel t ,
 x_t – Google indeksi väärtus ajahetkel t ,
 $\beta_{0h}; \beta_{ih}$ – baasmudeli parameetrid,
 $\alpha_{0h}; \alpha_{ih}$ – laiendatud mudeli (tegeliku aegrea osa) parameetrid,
 γ_{jh} – laiendatud mudeli (Google indeksi aegrea osa) parameetrid,
 h – prognoositava ajaperioodi järk ehk horisont,
 $i; j$ – parameetrite indeksid.

Prognoosimudelite valemite ülesehituse ülevaade on toodud lisas 2. Prognoositud aegride väärtused $y_{p\ t}$ sõltuvad suurel määral mudelite $M(0.h)$ parameetrite β (baasmudelid) ja $M(1.h)$ parameetrite α ja γ (laiendatud mudelid) väärtustest. Selleks, et leida mudelile kõige optimaalsem kordajate komplekt, leitakse kõigepealt tegeliku aegrea väärtuste y_t ja prognoositud aegrea väärtuste $y_{p\ t}$ vaheline ruutkeskmine viga RMSE.

$$(14) \quad RMSE = \sqrt{\frac{\sum_{t=t_1}^{t_2} (y_t - y_{p\ t})^2}{t_2 - t_1}}$$

kus $y_{p\ t}$ – prognoositud aegrea väärtus ajahetkel t ,
 y_t – tegeliku aegrea väärtus ajahetkel t ,
 t_1 – võrreldavate aegride esimene ühine ajahetk,
 t_2 – võrreldavate aegride viimane ühine ajahetk.

Seejärel leitakse mudelile kõige optimaalsem kordajate β (baasmudelid) või α ja γ (laiendatud mudelid) väärtuste komplekt kasutades tabelarvutustarkvara Microsoft Excel „Solver“ (*GRG Nonlineaar, multistart*) funktsiooni nii, et tegeliku aegrea ja prognoositud aegrea vaheline ruutkeskmine viga RMSE oleks minimaalne. Teisisõnu surutakse prognoosiva mudeli aegrea graafik võimalikult sarnaseks tegeliku aegrea graafikuga võrreldes. Taoline lähenemine, kus parameetrite hinnangud leitakse minimeerides mudeli samm-sammulist prognoosiviga (*multistep prediction error minimization method* – PEM), erineb mõnevõrra klassikalisest vähimruutude meetodist, kus parameetrid leitakse kogu andmestikku kasutades. Antud lähenemine on lähedalt seotud süsteemiteooriaga, mis kasutab otseselt vaadeldud süsteemi dünaamilist struktuuri. Pikemalt on võimalik antud lähenemisega tutvuda näiteks Petre Stoica, Arye Nehorai (1989) *On multistep*

prediction error methods for time series models töödes ja muudes prognoosivea minimeerimisega (lühidalt: *prediction error method* PEM) seotud käsitlustes (Pelckmans 2012).

Mineviku prognoosimisel kasutatakse kordajate optimeerimisel väljaarvutatud $y_{p\ t}$ väärtuseid:

- kvartali detailsusega prognoosimudelitel puhul 36 ajahetke (9 aastat),
- kuu detailsusega prognoosimudelitel puhul 108 ajahetke (9 aastat).

Tuleviku prognoosimisel kasutatakse kordajate optimeerimisel väljaarvutatud $y_{p\ t}$ väärtuseid:

- kvartali detailsusega prognoosimudelitel puhul 40 ajahetke (10 aastat),
- kuu detailsusega prognoosimudelitel puhul 120 ajahetke (10 aastat).

Kuna eelpool toodud prognoosimise loogika ülesehituse üldistav kirjeldus võib olla esmapilgul raskesti jälgitav, näidatakse järgnevalt prognoosimise loogika ülesehitust kinnisvara ostu-müügitehingute arvu mineviku (*ex post*) viis-sammu-ette prognoosimise näitel.

Kinnisvara ostu-müügitehingute arvu mineviku prognoosimise baasajahetk on neljas kvartal 2015. Selleks hetkeks on teada tegelikud kinnisvara ostu-müügitehingute arvud alates esimene kvartal 2004 (antud töö raames) kuni teine kvartal 2015 tegelikud andmed – moodustub antud näite mineviku prognoosimise (*ex post*) prognoosimise segment. Samuti on teada Google otsingute märksõnade arv alates jaanuar 2004 kuni september 2015. Antud algandmete alusel luuakse kinnisvara Google indeks esimene kvartal 2004 kuni 3 kvartal 2015. Rohkem andmeid teada ei ole, mille alusel prognoosimist teostada.

Etteruttavalt võib muidugi mainida, et lisaks eelnevale on (jaanuar 2017 seisuga) teada ka kinnisvara ostu-müügitehingute tegelikud arvud kolmas kvartal 2015 kuni kolmas kvartal 2017 – nendest moodustub testsegment.

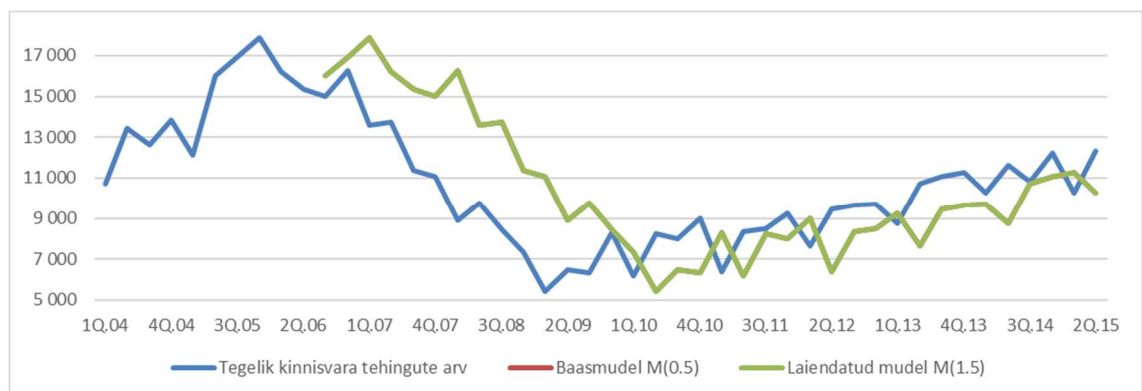
Eelnevat arvesse võttes, et neljandas kvartalis 2015 prognoosida näiteks kolmanda kvartali 2016 andmeid peab teostama teadaolevate algandmete tuginedes

prognoosimise viis-sammu-ette. Horisont = 5 ja kasutame mudeleid M(0.5) ja M(1.5) valemist 10 ja 11.

Tabel 2. Mudelite M(0.5) ja M(1.5) parameetrite β või α ja γ vaikimisi väärtused enne optimeerimist

Baasmudel M(0.5)		Laiendatud mudel M(1.5)	
Mudeli parameetrid		Mudeli parameetrid	
β_{45}	0,000000000	α_{45}	0,000000000
β_{35}	0,000000000	α_{35}	0,000000000
β_{25}	0,000000000	α_{25}	0,000000000
β_{15}	1,000000000	α_{15}	1,000000000
β_{05}	0,000000000	α_{05}	0,000000000
		γ_{45}	0,000000000
		γ_{35}	0,000000000
		γ_{25}	0,000000000
		γ_{15}	0,000000000
		γ_{05}	0,000000000

Sealjuures tuleb leida mudelitele kõige optimaalsemad kordajate β või α ja γ väärtused. Enne kordajate optimeerimist sisestatakse β või α ja γ vaikimisi väärtused vastavalt tabelile 2. Ehk optimeerimata viis-sammu-ette prognoosimisel võetakse arvesse ainult viimase teadaoleva kvartali andmed. Kordajate vaikimisi väärtuste komplekti valiku tulemusel ühtib baasmudeli ja laiendatud mudeli graafik Joonisel 22, nihutades tegeliku kinnisvara ostu-müügilepingute arvu graafikut lihtsalt viie kvartali võrra.



Joonis 22. Kinnisvara ostu-müügitehingute tegeliku arvu ja mudelite M(0.5) ja M(1.5) graafikud kordajate β või α ja γ vaikimisi väärtuste juures, tk

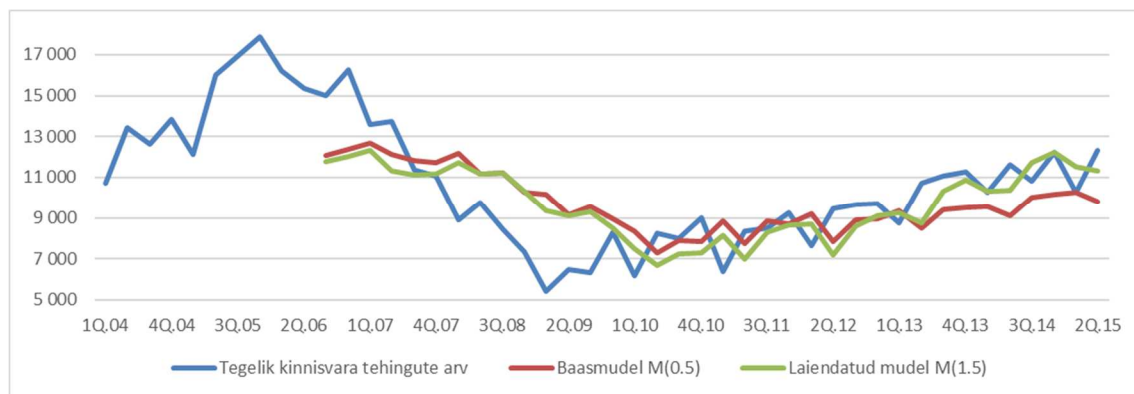
Baasmudelite kordajate β optimeerimiseks arvutatakse esmalt teise kvartali 2015 prognoos (kasutades teise kvartali 2013 kuni esimese kvartali 2014 tegelikke andmeid) ja arvutatakse teise kvartali 2015 prognoosi ning tegeliku kinnisvara ostu-müügilepinute arvu vahe ruut. Ajas tagasi minnes arvutatakse teisalt esimese kvartali 2015 prognoos (kasutades esimese kvartali 2013 kuni neljanda kvartali 2013 tegelikke andmeid) ja arvutatakse esimese kvartali 2015 prognoosi ja tegeliku kinnisvara ostu-müügilepinute arvu vaheline vahe ruut. Ja niimoodi ajas tagasi minnes kuni teostatakse kolmanda kvartali 2006 (kasutades kolmanda kvartali 2004 kuni teise kvartali 2005 tegelikke andmeid) prognoos ja arvutatakse kolmanda kvartali 2006 prognoosi ja tegeliku kinnisvara ostu-müügilepinute arvu vaheline vahe ruut. Saadakse 36 kvartali prognoosi ja tegeliku vahe ruutude avaldist. Nendest 36 avaldisest võetakse keskmine ja seejärel ruutjuur ehk kokkuvõtvalt kasutatakse ruutkestmise vea RMSE valemit 14.

Tabel 3. Mudelite M(0.5) ja M(1.5) parameetrite β või α ja γ väärtused peale optimeerimist

Baasmudel M(0.5)		Laiendatud mudel M(1.5)	
Mudeli parameetrid		Mudeli parameetrid	
β_{45}	0,000000000	α_{45}	0,000000000
β_{35}	0,000000000	α_{35}	0,000000000
β_{25}	0,000000000	α_{25}	0,000037010
β_{15}	0,468840611	α_{15}	0,643640523
β_{05}	2,110515237	α_{05}	0,692264766
		γ_{45}	0,000000000
		γ_{35}	0,000000000
		γ_{25}	0,000000000
		γ_{15}	0,417967625
		γ_{05}	0,000000000

Seejärel leitakse mudelile kõige optimaalsem kordajate β (baasmudelid) või α ja γ (laiendatud mudelid) väärtuste komplekt kasutades tabelarvutustarkvara Microsoft Excel „Solver“ (*GRG Nonlineaar, multistart*) funktsiooni nii, et tegeliku aegrea ja prognoositud aegrea vaheline ruutkeskmine viga RMSE oleks minimaalne. Selliselt töötabki prognoosivigadel baseeruv parameetrite hindamine (*predcition error method* PEM). Tulemust näeb tabelist 3 ja jooniselt 23. Tabelist on näha, et kinnisvara viis-sammu-ette ostumüügitehingute arvu baasmudeli prognoosimisel kasutatakse ainult viimase

teadaoleva kvartali informatiivseid andmeid tõstes lõppväärtusi vabateguri abil. Seevastu laiendatud mudelil kasutatakse lisaks viimase teadaoleva kvartali andmetele ka eelviimase kvartali andmeid ning lisaks Google indeksi eelviimase kvartali andmeid. Samaväärset loogikat kasutatakse laiendatud mudeli puhul, kus lisaks tegelikele andmetele võetakse arvesse ka vastavaid Google indeksi aegrea andmeid.



Joonis 23. Kinnisvara ostu-müügitehingute tegeliku arvu ja mudelite M(0.5) ja M(1.5) graafikud kordajate β või α ja γ optimeeritud väärtuste juures, tk

Viis-sammu-ette kinnisvara ostumüügitehingute arvu mineviku prognoosimise baasmudeli ruutkeskmise viga võrreldes tegelike andmetega on 2019,57 tehingut ja laiendatud mudelil 1853,26 tehingut. Antud ruutkeskmiste vigade väärtuseid mingiks analüüsiks ei kasutata. Nagu juba eelnevast järeldub, kasutatakse ruutkeskmise vea arvutamist ainult mudelite optimaalse kordajate komplekti leidmiseks.

Peatüki algusesse tagasi tulles arvutatakse optimeeritud kordajate β väärtused ning valemit 10 kasutades välja kolmanda kvartali 2016 baasmudeli prognoos - 10682,51 tehingut (kasutades kolmanda kvartali 2014 kuni teise kvartali 2015 tegelike). Kordajate α ja γ väärtuseid ning valemit 11 kasutades arvutatakse välja kolmanda kvartali 2016 laiendatud mudeli prognoos - 12783,34 tehingut (kasutades kolmanda kvartali 2014 kuni teise kvartali 2015 tegelikke ning kolmanda kvartali 2014 kuni kolmanda kvartali 2015 Google indeksi andmeid). Neid kahte arvu võrreldakse testsegmentist teadaoleva tegeliku kolmanda kvartali 2016 kinnisvara ostu-müügitehingute arvuga – 12245 tehingut (vt ka tabel 4).

Tabel 4. Tegelik kinnisvara ostu-müügitehingute arv ja Mudelite M(0.5) ja M(1.5) prognoosid kolmanda kvartali jaoks 2016

Järk h	Kvartal	Kinnisvara ostu-müügilepingute arvu (mineviku) prognoos		
		Tegelik [tk]	Baasmudel M(0.h)	Laiendatud mudel M(1.h)
5	3kv.16	12245	10682,51	12783,34

Mõlema jaoks arvutatakse välja (absoluutne) suhteline viga APE (*Absolute Percentage Error*) (vt ka tabel 5). Suhtelise vea arvutamist APE kasutatakse vahetulemusena hilisemalt keskmise absoluutse suhtelise vea MAPE arvutamisel. Tabelis 5 on ära toodud kinnisvara ostu-müügilepingute arvu mineviku viis-sammu-ette prognoosimise (*ex post*) kolmanda kvartali 2016 suhtelise vea APE tulemused võrreldes tegeliku kolmanda kvartali kinnisvara ostu-müügilepingute arvuga.

Tabel 5. Kinnisvara ostu-müügitehingute tegeliku arvu ja Mudelite M(0.5) ja M(1.5) prognoosi vaheline suhteline viga

Järk h	Kvartal	APE		
		Baasmudel M(0.5)	Laiendatud mudel M(1.5)	Parendus M(0.5) - M(1.5)
5	3kv.16	12,76%	4,40%	8,36%

Tabelist järeldub, et laiendatud mudeli väljapakutud prognoos on suutnud vähendada viga 8,36% punkti võrra, ehk laiendatud mudeli prognoos on lähem tegelikule.

Tabel 6. Individuaalsete kordajate ja kordajate komplektide arv

Mineviku prognoosimine	Näitajate arv	Kordajate komplektide arv	Kordajate arv		Kokku
			β	α ja γ	
Kvartali detailsusega mudelid	2	5	5	10	150
Kuu detailsusega mudelid	2	15	13	27	1200
Tuleviku prognoosimine					
Kvartali detailsusega mudelid	2	5	5	10	150
Kuu detailsusega mudelid	2	15	13	27	1200

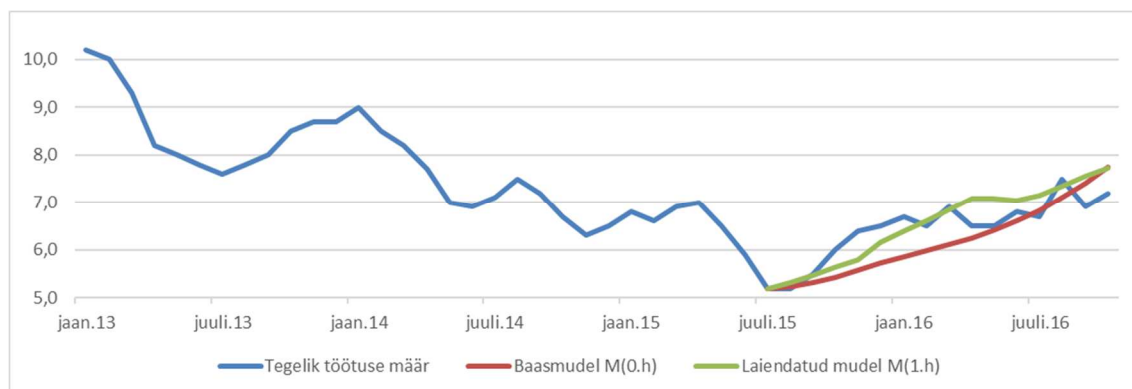
Hinnatud kordajate koguarv üle kõigi mudelite: 2700

Eelnevalt on kirjeldatud mineviku, kvartali detailsusega mudelite viis-sammu-ette prognoosimise ülesehitust. Täpselt sama lähenemist kasutatakse nii tuleviku, h-sammu-ette, kui ka kuu detailsusega prognoosimise läbiviimisel. Igal prognoositud

üksikväärtusel on ainuomane kordajate β (baasmudelid) või α ja γ (laiendatud mudelid) väärtuste komplekt. Kokku on antud töö raames välja arvutatud 80 β ja 80 α ja γ väärtuste komplekti 2700 individuaalse kordajaga, et välja pakkuda 80 baasmudeli üksikprognosi ja 80 laiendatud mudeli üksikprognosi (vt ka tabel 6).

3.3. Mineviku prognoosimine

Kasutades eelpoolkirjeldatud metoodikat ja valemeid 12, 13 ja 14 arvutatakse välja töötuse määra prognoos iga vaatluse all oleva viieteistkümne ajahetke ehk kuu (kaks mineviku + üks oleviku + kaksteist tuleviku) jaoks. Prognoose võrreldakse tegeliku töötuse määra andmetega kasutades valemit 9. Töötuse määra prognoosid ja prognoosimise viga (MAPE) tegeliku töötuse määra suhtes on toodud tabelis 7 ja joonisel 24.



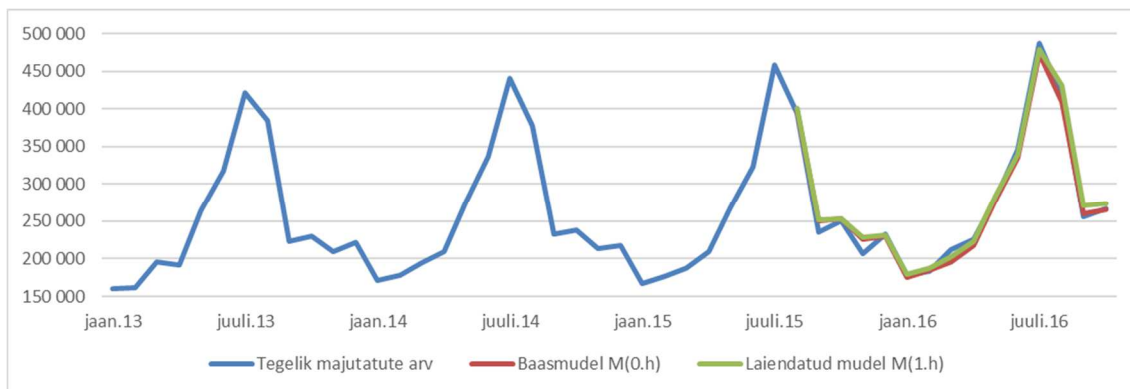
Joonis 24. Tegelik töötuse määr ja töötuse määra mineviku prognoosid, %

Tabelist 7 ja jooniselt 24 võib märgata, et nii baasmudeli kui ka laiendatud mudeli poolt prognoositud ajahetkede arväärtuste vead võrreldes tegeliku töötuse määra arväärtustega ei ole katastroofilised, jäädes suurusjärgus 10% piiridesse. Samas on baasmudeli maksimaalne viga 12,82% ja minimaalne 0,58% (seda küll kohe ajaperioodi esimesel ajahetkel). Laiendatud mudeli vead häälbivad aga 0,71% ja 9,62% vahemikus. Sellest tulenevalt on ka laiendatud mudeli keskmine absoluutne suhteline viga 1,53% võrra väikesem. Seega võib väita, et Interneti-otsingu andmete kaasamine töötuse määra prognoosimisel omab mõtet.

Tabel 7. Töötuse määra mineviku prognoosid ja prognoosimise viga (MAPE)

		Töötuse määr mineviku prognoosid			APE		
Järk h	Kuu	Tegelik	Baasmudel M(0.h)	Laiendatud mudel M(1.h)	Baasmudel M(0.h)	Laiendatud mudel M(1.h)	Parendus M(0.h) - M(1.h)
1	aug.15	5,20%	5,23%	5,32%	0,58%	2,33%	-1,75%
2	sept.15	5,50%	5,31%	5,46%	3,39%	0,71%	2,67%
3	okt.15	6,00%	5,43%	5,64%	9,49%	6,04%	3,45%
4	nov.15	6,40%	5,58%	5,78%	12,82%	9,62%	3,20%
5	dets.15	6,50%	5,72%	6,16%	11,99%	5,24%	6,75%
6	jaan.16	6,70%	5,86%	6,40%	12,54%	4,46%	8,08%
7	veebr.16	6,50%	5,99%	6,61%	7,91%	1,63%	6,28%
8	märts.16	6,90%	6,11%	6,84%	11,42%	0,90%	10,52%
9	apr.16	6,50%	6,25%	7,08%	3,87%	8,89%	-5,02%
10	mai.16	6,50%	6,41%	7,07%	1,37%	8,80%	-7,42%
11	juuni.16	6,80%	6,61%	7,04%	2,83%	3,51%	-0,68%
12	juuli.16	6,70%	6,83%	7,14%	2,01%	6,57%	-4,57%
13	aug.16	7,50%	7,10%	7,35%	5,31%	2,03%	3,27%
14	sept.16	6,90%	7,41%	7,56%	7,38%	9,52%	-2,14%
15	okt.16	7,20%	7,75%	7,74%	7,68%	7,44%	0,24%
	MAPE	6,52%	6,24%	6,61%	6,71%	5,18%	1,53%

Kasutades eelpoolkirjeldatud metoodikat ja valemeid 12, 13 ja 14 arvutatakse järgnevalt välja majutatute arvu prognoos iga vaatluse all oleva viieteistkümne ajahetke ehk kuu (kaks mineviku + üks oleviku + kaksteist tuleviku). Prognoose võrreldakse taas tegeliku majutatute arvu andmetega kasutades valemit 9. Majutatute arvu prognoosid ja prognoosimise viga (MAPE) tegeliku majutatute arvu suhtes on toodud tabelis 8 ja joonisel 25.



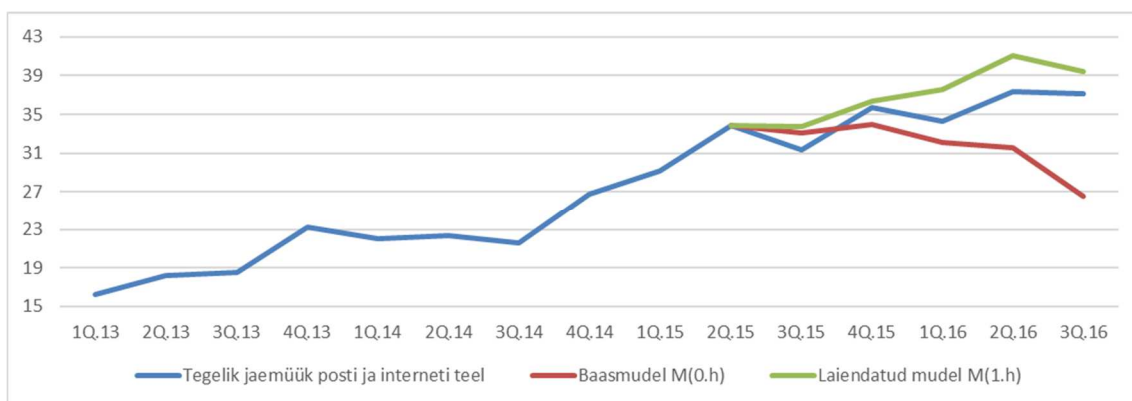
Joonis 25. Tegelik majutatute arv ja majutatute arvu mineviku prognoosid, tk

Tabelist 8 ja jooniselt 25 on näha, et mõlemad prognoosid järgivad ootamatult hästi testsegmendi andmeid. Sealjuures tuleb märkida, et laiendatud mudel ei vähenda keskmiselt prognoosimise viga baasmudeliga võrreldes ehk Google otsingute andmete kaasamine prognoosimudelitesse ei oma erilist mõtet. Käesolevas töös ei leitud optimaalset komplekti Google märksõnu ja nende arvu, et laiendatud mudeli keskmine absoluutne suhteline viga (MAPE) võrreldes tegeliku majutatud arvuga oleks väikesem baasmudeli omast. Põhjus võib leiduda selles, et baasmudeli keskmine absoluutne suhteline viga (MAPE) võrreldes tegeliku majutatud arvuga on juba piisavalt madal (2,87%). Sellele vaatamata pakutakse majutatute arvu kohta välja ka tuleviku prognoosid ka laiendatud mudelist.

Tabel 8. Majutatute arvu mineviku prognoosid ja prognoosimise viga (MAPE)

Järk h	Kuu	Majutatute arvu mineviku prognoosid			APE		
		Tegelik [tk]	Baasmudel M(0.h)	Laiendatud mudel M(1.h)	Baasmudel M(0.h)	Laiendatud mudel M(1.h)	Parendus M(0.h) - M(1.h)
1	aug.15	393 539	398 332	401 593	1,22%	2,05%	-0,83%
2	sept.15	235 271	249 967	252 511	6,25%	7,33%	-1,08%
3	okt.15	250 853	252 689	253 471	0,73%	1,04%	-0,31%
4	nov.15	207 006	226 054	228 421	9,20%	10,35%	-1,14%
5	dets.15	233 182	230 278	230 955	1,25%	0,96%	0,29%
6	jaan.16	178 204	175 049	179 842	1,77%	0,92%	0,85%
7	veebr.16	183 305	184 365	187 435	0,58%	2,25%	-1,67%
8	märts.16	212 442	196 078	202 279	7,70%	4,78%	2,92%
9	apr.16	226 369	217 921	223 125	3,73%	1,43%	2,30%
10	mai.16	279 540	278 116	282 839	0,51%	1,18%	-0,67%
11	juuni.16	345 630	335 046	339 529	3,06%	1,77%	1,30%
12	juuli.16	487 681	473 587	479 696	2,89%	1,64%	1,25%
13	aug.16	417 242	408 717	431 204	2,04%	3,35%	-1,30%
14	sept.16	256 653	259 964	270 557	1,29%	5,42%	-4,13%
15	okt.16	267 297	265 023	273 369	0,85%	2,27%	-1,42%
	MAPE	278 281	276 746	282 455	2,87%	3,11%	-0,24%

Kasutades eelpoolkirjeldatud metoodikat ja valemeid 10, 11 ja 14 arvutatakse välja jaemüügi posti või Interneti teel prognoos iga vaatluse all oleva viie ajahetke ehk kvartali (üks mineviku + üks oleviku + kolm tuleviku) jaoks. Prognoose võrreldakse tegeliku jaemüügi posti või Interneti teel andmetega kasutades valemit 9. Jaemüügi posti või Interneti teel prognoosid ja prognoosimise viga (MAPE) tegeliku jaemüügi posti või Interneti teel suhtes on toodud tabelis 9 ja joonisel 26.



Joonis 26. Tegelik jaemüügi posti või Interneti teel ja jaemüügi posti või Interneti teel mineviku prognoosid, mln. eurot

Tabelit 9 ja joonist 26 analüüsid järele, et laiendatud mudel on suutnud järgida tegelike andmete trendi, baasmudel aga mitte. Tegemist on üllatava tulemusega just baasmudelit silmas pidades, sest vaadates ajaloolisi tegelike andmeid, siis jaemüügile posti või Interneti teel nii jõulist langustrendi ei julgeks ennustada. Seevastu laiendatud mudel järgib hästi tegelike andmete trendi. Iseseisvalt ei ole laiendatud mudel muidugi kuigi täpne (võrreldes just teiste töös kasutatavate näitajatega), andes üks-samm-ette ja kolm-sammu-ette baasmudeli prognoosist halvemaid tulemusi. Kuna aga baasmudeli neljanda ja viienda-samm-ette prognoosi viga võrreldes tegelike andmetega on väga suur, siis kokkuvõtvalt suudab laiendatud mudel edastada baasmudelit keskmiselt 5,15% võrra ehk Interneti-otsingu andmete kaasamine prognoosimisele omab mõtet.

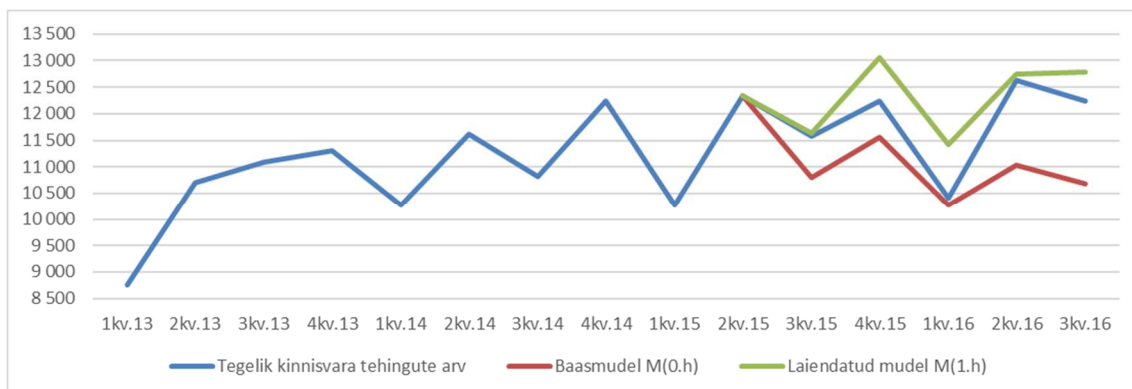
Viimasena arvutatakse välja kasutades eelpoolkirjeldatud metoodikat ja valemeid 10, 11 ja 14 kinnisvara ostu-müügitehingute arvu prognoos iga vaatluse all oleva viie ajahetke ehk kvartali (üks mineviku + üks oleviku + kolm tuleviku) jaoks. Analoogselt eestooduga võrreldakse prognoose tegeliku kinnisvara ostu-müügitehingute arvu andmetega kasutades valemit 9.

Tabel 9. Jaemüügi posti või Interneti teel mineviku prognoosid ja prognoosimise viga (MAPE)

		Jaemüük posti ja Interneti teel mineviku prognoos			APE		
Järk h	Kvartal	Tegelik [miljonit eurot]	Baasmudel M(0.h)	Laiendatud Mudel M(1.h)	Baasmudel M(0.h)	Laiendatud mudel M(1.h)	Parendus M(0.h) - M(1.h)
1	3kv.15	33,80	33,12	33,75	5,81%	7,84%	-2,03%
2	4kv.15	31,30	33,93	36,41	4,95%	2,00%	2,95%
3	1kv.16	35,70	32,09	37,55	6,43%	9,47%	-3,04%
4	2kv.16	34,30	31,56	41,03	15,61%	9,71%	5,90%
5	3kv.16	37,40	26,57	39,47	28,37%	6,39%	21,98%
	MAPE	34,50	31,46	37,64	12,23%	7,08%	5,15%

Tabel 10. Kinnisvara ostu-müügitehingute arvu mineviku prognoosid ja prognoosimise viga (MAPE)

		Kinnisvara ostu-müügilepingute arvu mineviku prognoos			APE		
Järk h	Kvartal	Tegelik [tk]	Baasmudel M(0.h)	Laiendatud mudel M(1.h)	Baasmudel M(0.h)	Laiendatud mudel M(1.h)	Parendus M(0.h) - M(1.h)
1	3kv.15	11 581	10 809	11 632	6,67%	0,44%	6,23%
2	4kv.15	12 245	11 551	13 049	5,67%	6,56%	-0,89%
3	1kv.16	10 403	10 273	11 429	1,25%	9,86%	-8,62%
4	2kv.16	12 628	11 033	12 737	12,63%	0,86%	11,76%
5	3kv.16	12 245	10 683	12 783	12,76%	4,40%	8,36%
	MAPE	11 820	10 870	12 326	7,79%	4,42%	3,37%

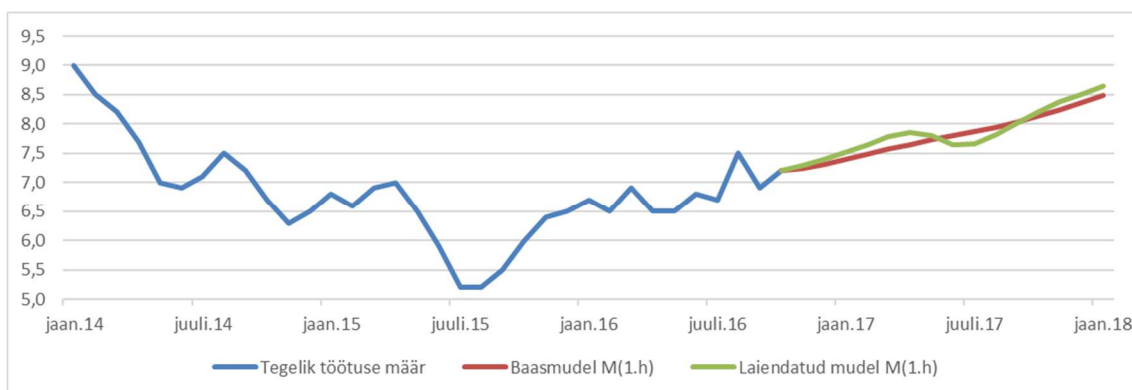


Joonis 27. Tegelik kinnisvara ostu-müügitehingute arv ja kinnisvara ostu-müügitehingute arvu mineviku prognoosid, tk

Vaadates tabelit 10 ja joonist 27 näeme, et nii baasmudel kui ka laiendatud mudel annavad sarnaseid tulemusi, suutes oma prognoosides järgida tegeliku kinnisvara ostu-müügilepingute arvu muutuste trendi. Kuigi laiendatud mudel suudab kokkuvõttes edestada baasmudelit 3,37% võrra, siis lähemalt vaadates on ka laiendatud mudeli prognooside vead kohati (eriti kolm-sammu-ette – 8,62%) suhteliselt nigelad. Siiski kokkuvõttes näib, et Google otsingute andmete kaasamine prognoosimudelitesse omab mõtet.

3.4. Tuleviku prognoosimine

Kasutades eelpoolkirjeldatud metoodikat ja valemeid 12, 13 ja 14 arvutatakse välja töötuse määra prognoos iga vaatluse all oleva ajahetke (2+1+12) jaoks. Töötuse määra tuleviku prognoosid on toodud tabelis 11 ja joonisel 28.



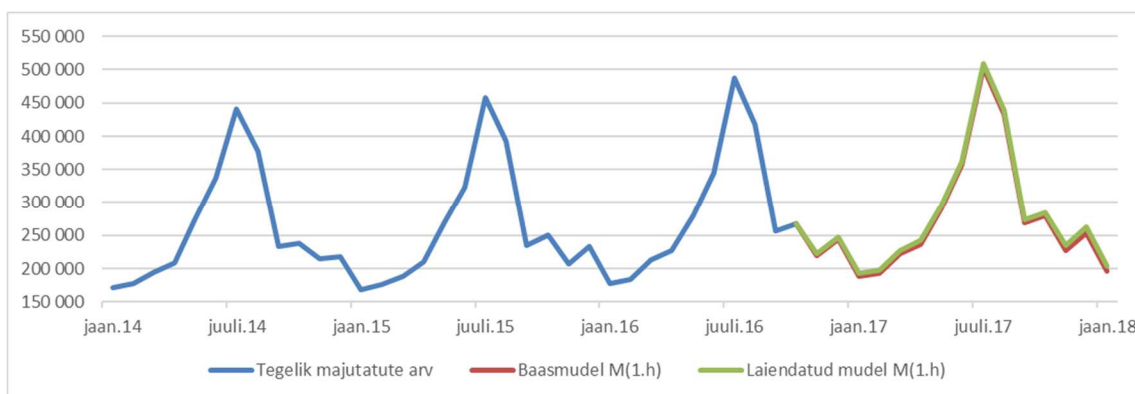
Joonis 28. Tegelik töötuse määr ja töötuse määra tuleviku prognoosid, %

Tabel 11. Töötuse määra tuleviku prognoosid

Järk h	Kuu	Töötuse määra tuleviku prognoos [%]	
		Baasmudel M(0.h)	Laiendatud mudel M(1.h)
1	nov.16	7,24%	7,29%
2	dets.16	7,31%	7,40%
3	jaan.17	7,40%	7,53%
4	veebr.17	7,49%	7,64%
5	märts.17	7,57%	7,78%
6	apr.17	7,65%	7,85%
7	mai.17	7,73%	7,80%
8	juuni.17	7,80%	7,64%
9	juuli.17	7,87%	7,67%
10	aug.17	7,94%	7,82%
11	sept.17	8,03%	8,01%
12	okt.17	8,13%	8,21%
13	nov.17	8,24%	8,39%
14	dets.17	8,36%	8,50%
15	jaan.18	8,49%	8,65%

Analoogselt töötuse määra mineviku prognoosimise tulemusega, prognoosivad ka tuleviku prognoosimudelid töötuse määrale kerget tõusu. Laiendatud mudeli graafikul on selgesti märgatav juulikuine kerge madalseis. Baasmudeli prognoos tundub seevastu suhteliselt lineaarselt tõusev. Arvestades seda, et mineviku prognoosimisel esines maksimaalseid vigu tegeliku suhtes baasmudelitel kuni 13% ja laiendatud mudelitel kuni 10%, siis usaldusväärsemate prognooside andmiseks tuleks leida usaldusväärsemad prognoosimudelid võrreldes käesolevas töös kasutatud mudelitele. Kuna tegeliku töötuse määra ja töötuse määra Google indeksi graafiline analüüs näitas, et töötuse määra Google indeks ennetab tegeliku töötuse määra trende keskeltläbi kümme kuud, siis on mõeldav töötuse määra Google indeksi väärtustesse või veel parem laiendatud mudeli Google indeksi osasse viia sisse lisa viitaeg kümme kuud. Samuti on mõeldav lihtsalt suurendada laiendatud mudeli Google indeksi komponentide arvu kümne võrra, kokku 24 komponendini. Teisest küljest võtavad töös kasutatud mudelid juba arvesse töötuse määra kahetestkümne kuu vanuseid andmeid, seega lisa viitaja sissetoomine ei pruugi täiendavat efekti luua.

Kasutades eelpoolkirjeldatud metoodikat ja valemeid 12, 13 ja 14 arvutatakse välja majutatute arvu prognoos iga vaatluse all oleva ajahetke (2+1+12) jaoks. Majutatute arvu tuleviku prognoosid on toodud tabelis 12 ja joonisel 29. Kuigi antud töö eesmärgiks ei olnud leida vaatluse all olevate näitajate jaoks parimaid prognoosimudeleid, siis majutatute arvu mineviku prognoosimise näitaja varal võib öelda, et nii baasmudel kui ka laiendatud mudel annavad sarnaseid ja väikese veaga tulemusi. Ka tuleviku prognoosimisel ei erine baasmudeli ja laiendatud mudeli prognoos teineteisest kuigivõrd. Siiski võib graafikutelt märgata ajaloolise kerge tõusuga sesoonse trendi jätkumist.

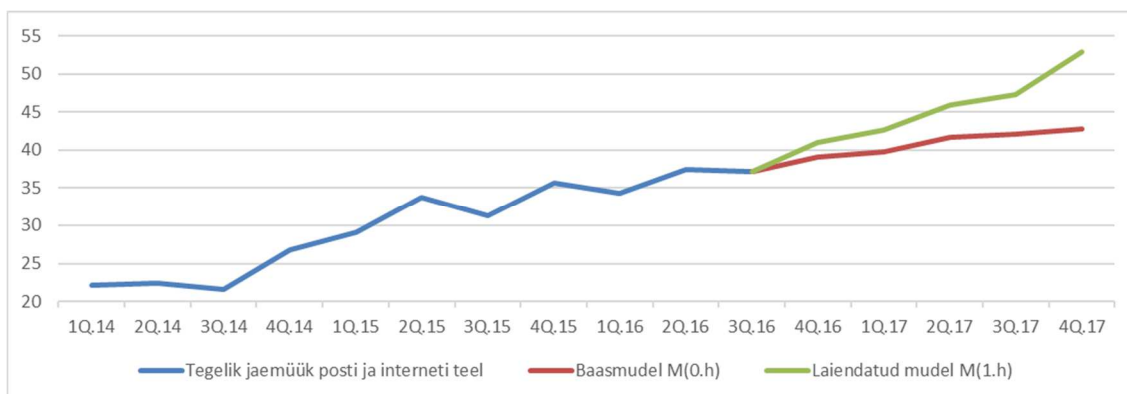


Joonis 29. Tegelik majutatute arv ja majutatute arvu tuleviku prognoosid, tk

Tabel 12. Majutatute arvu tuleviku prognoosid, tk

Järk h	Kuu	Majutatute arvu tuleviku prognoos [tk]	
		Baasmudel M(0.h)	Laiendatud mudel M(1.h)
1	nov.16	218 864	223 033
2	dets.16	240 626	240 088
3	jaan.17	188 009	192 242
4	veebr.17	192 296	197 188
5	märts.17	222 741	227 036
6	apr.17	236 534	241 834
7	mai.17	290 027	295 484
8	juuni.17	358 067	362 380
9	juuli.17	502 720	508 978
10	aug.17	432 800	439 400
11	sept.17	269 699	273 237
12	okt.17	279 617	284 751
13	nov.17	227 301	235 397
14	dets.17	253 841	262 380
15	jaan.18	195 625	203 630

Kasutades eelpoolkirjeldatud metoodikat ja valemeid 10, 11 ja 14 arvutatakse välja jaemüügi posti või Interneti teel prognoos iga vaatluse all oleva ajahetke (1+1+3) jaoks. Jaemüügi posti või Interneti teel tuleviku prognoosid on toodud tabelis 13 ja joonisel 30.



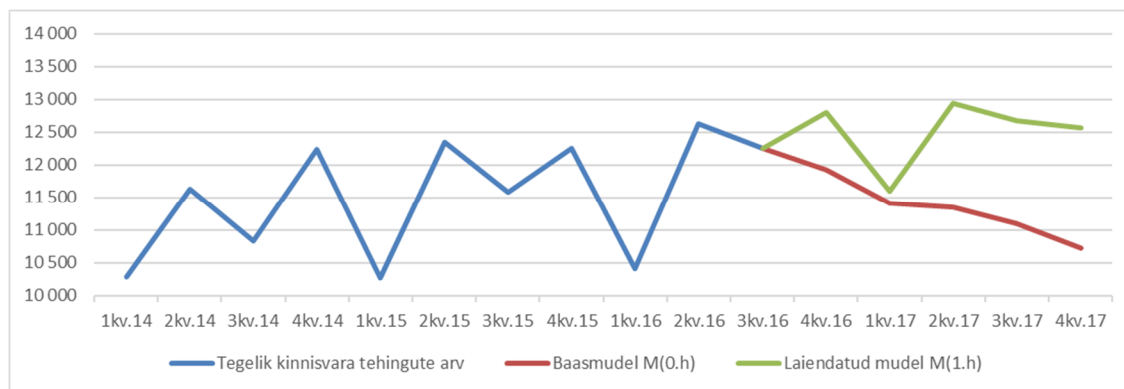
Joonis 30. Tegelik jaemüük posti või Interneti teel ja jaemüügi posti või Interneti teel tuleviku prognoosid, mln. eurot

Tabelist 13 ja jooniselt 30 võib näha, et nii baasmudel kui ka laiendatud mudel prognoosib jaemüügi posti ja Interneti teel kasvu lähiaastal. Erinevus tundubki olema kasvu kiiruses. Toetudes mineviku prognoosimise tulemustele on üks-kuni kolm-sammu-ette prognoosimise tulemused väiksema veaga ja seega usaldusväärsemad. Samas võib neljas (jõulueelne) kvartal üllatada.

Tabel 13. Jaemüügi posti või Interneti teel tuleviku prognoosid, mln. eurot

Järk h	Kvartal	Jaemüük posti või Interneti teel tuleviku prognoos [miljonit eurot]	
		Baasmudel M(0.h)	Laiendatud mudel M(1.h)
1	4kv.16	39,11	41,03
2	1kv.17	39,73	42,67
3	2kv.17	40,42	45,93
4	3kv.17	42,14	47,33
5	4kv.17	42,73	52,91

Kasutades eelpoolkirjeldatud metoodikat ja valemeid 10, 11 ja 14 arvutatakse välja kinnisvara ostu-müügitehingute arvu prognoos iga vaatluse all oleva ajahetke (1+1+3) jaoks. Kinnisvara ostu-müügitehingute arvu tuleviku prognoosid on toodud tabelis 14 ja joonisel 31.



Joonis 31. Tegelik kinnisvara ostu-müügitehingute arv ja kinnisvara ostu-müügitehingute arvu tuleviku prognoosid, tk

Võttes arvesse mineviku prognoosimise järeldust, et Google otsingute andmete kaasamine prognoosimudelitesse omab mõtet, on oodata kinnisvara ostu-müügitehingute arvu graafiku käitumist vastavalt laiendatud mudeli prognoosile.

Tabel 14. Kinnisvara ostu-müügitehingute arvu tuleviku prognoosid, tk

Järk h	Kvartal	Kinnisvara ostu-müügitehingute arvu tuleviku prognoos [tk]	
		Baasmudel M(0.h)	Laiendatud mudel M(1.h)
1	4kv.16	11 919	12 799
2	1kv.17	11 398	11 600
3	2kv.17	11 349	12 932
4	3kv.17	11 094	12 665
5	4kv.17	10 728	12 560

Kokkuvõttes näeme, et Google otsingumootori andmete põhjal loodud indeksite kaasamine prognoosimudelitesse aitab tõepoolest kaasa prognooside täpsuse tõusule mitmel juhul. Osadel juhtudel, mis on niigi hästi prognoositavad, näiteks tugeva sesoonsusega majutatute arv, ei anna lisainformatsiooni kaasamine, vähemalt mitte stabiilsel perioodil, täiendavat efekti. Teistel juhtudel, näiteks töötuse määra puhul, näib Google otsingute märksõnade sageduse andmete kaasamine mudeli prognoosivõimet tõstvat. Näitajate täiendavaks analüüsiks on võimalik läbi viia trendipõhine või lühiajaline prognoosimine. Teisisõnu valitakse sobiv analüüsi alguspunkt erinevalt käesolevas töös kasutatud 1. jaanuarist 2004.

Tabel 15. Töötuse määra prognoosimise vea (MAPE) võrdlus Soome ja Ameerika Ühendriikide töötuse määra prognoosimise vigadega

	Tuhkuri 2014 (Soome)			Tuhkuri 2016 (Ameerika Ühendriigid)			Töötuse määr (Eesti)		
Järk h	Baasmudel M(0.h)	Laiendatud mudel M(1.h)	Parendus M(0.h) - - M(1.h)	Baasmudel M(0.h)	Laiendatud mudel M(1.h)	Parendus M(0.h) - - M(1.h)	Baasmudel M(0.h)	Laiendatud mudel M(1.h)	Parendus M(0.h) - - M(1.h)
1	7,80%	7,00%	0,80%	4,58%	4,38%	0,20%	0,58%	2,33%	-1,75%
2	9,30%	7,70%	1,60%	7,57%	7,01%	0,56%	3,39%	0,71%	2,67%
3	10,50%	7,00%	3,50%	9,48%	9,85%	-0,37%	9,49%	6,04%	3,45%
4	11,10%	6,70%	4,40%	10,40%	11,06%	-0,66%	12,82%	9,62%	3,20%
5	11,30%	7,70%	3,60%	11,10%	13,02%	-1,92%	11,99%	5,24%	6,75%
6	11,30%	8,40%	2,90%	11,96%	13,54%	-1,58%	12,54%	4,46%	8,08%
7	11,40%	9,00%	2,40%	13,40%	12,07%	1,33%	7,91%	1,63%	6,28%
MAPE	10,39%	7,64%	2,74%	9,78%	10,13%	-0,35%	8,39%	4,29%	4,10%

Allikas: (Tuhkuri 2014 ja 2016)

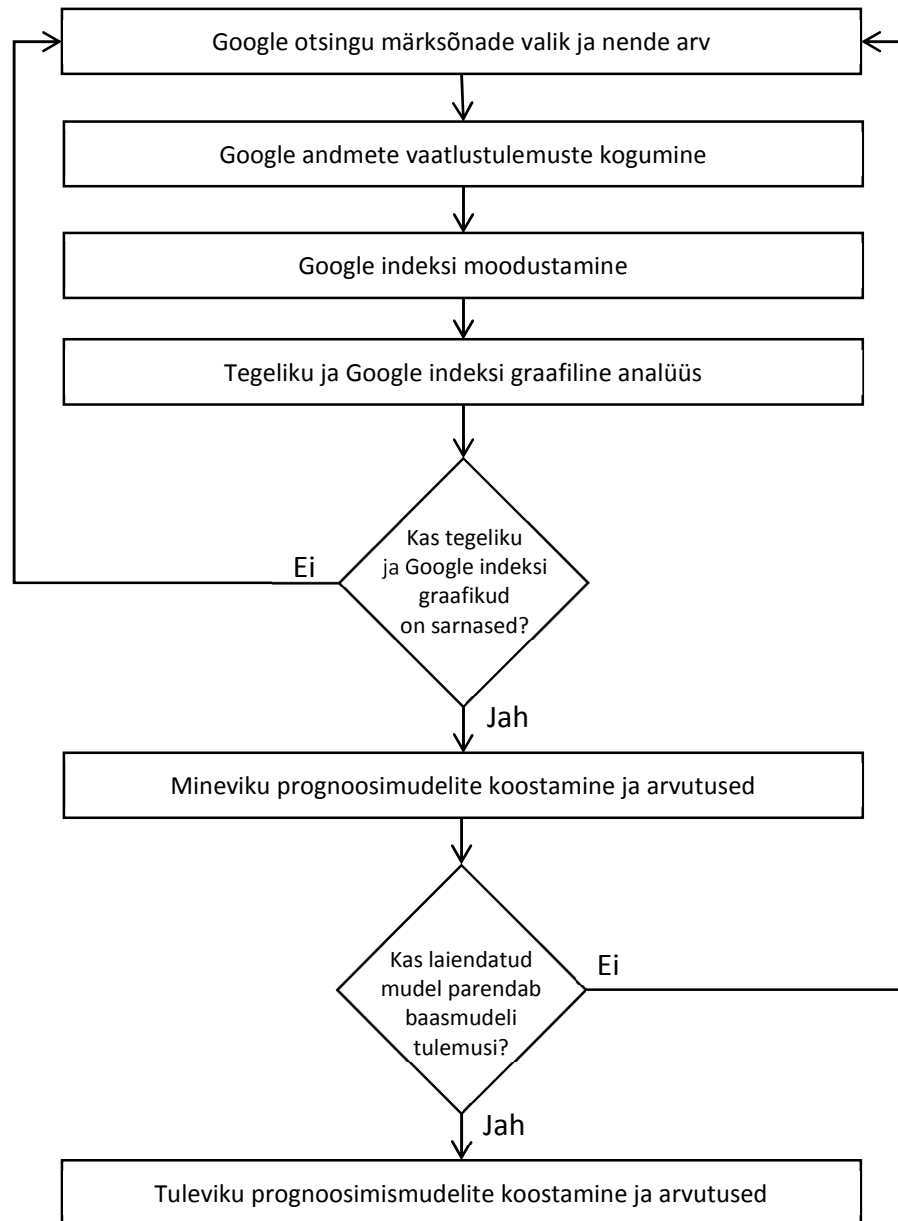
Töö empiirilise analüüsi läbiviimisel neid võimalusi ka testiti, kuid töö selguse ja analüüsi erinevate versioonide vältimiseks, trendipõhise ja lühiajalise prognoosimise analüüsi tulemusi käesolevas töös välja ei toodud.

Kuigi käesoleva töö ja Tuhkuri 2014 ja 2016 tööde tulemused (vt ka lisa 3) tunduvad esmapilgul võrreldavad, tuleb siiski ära märkida, et tulemused ei ole üks-ühele võrreldavad just seetõttu, et töödes on kasutatud erinevaid mudeleid ja erinevaid prognoosimise meetodeid (näiteks staatiline versus dünaamiline prognoosimine). Üks-ühele võrreldavate tulemuste saamiseks on vajalik kasutada samu mudeleid ja meetodeid, asendades lihtsalt Soome ja Ameerika Ühendriikide kohta käivad näitajate statistilised andmed Eesti näitajatega. Samuti peaksid olema kasutatud Google otsingute märksõnad ja Google indeksi moodustamise alused võrreldavad ning võrdluse saaks läbi viia ainult töötuse määra kohta. Käesoleva töö autori isikliku panuse suurendamiseks ja teiste autorite kordamise vältimiseks ei ole seda teed mindud. Sellele vaatamata on tabelis 15 välja toodud käesoleva töö ja Tuhkuri 2014 ja 2016 tööde tulemused töötuse määra kohta. Tabelist on näha, et käesoleva töö prognoosimise mudelid ja meetodit on suutnud näidata üldjuhul väiksemaid absoluutseid keskmise vea näitajaid kui Tuhkuri töödes. Sellest tulenevalt edestab käesoleva töö töötuse määra keskmine absoluutne keskmine viga Soome töötuse määra töö keskmist absoluutset keskmist viga 1,36% võrra ja Ameerika Ühendriikide töötuse määra töö keskmist absoluutset keskmist viga koguni 4,45% võrra.

Üldjoontes on käesolevas töös leitud tulemused kooskõlas varasemate uuringutega vaatamata tõsiasjale, et tehtud tööde fookus ja kasutatud meetodid on erinevad. Edaspidises uurimistöös tasuks töös kasutatud mudeleid edasi arendada, katsetada erinevaid funktsionaalseid seoseid ning ka alternatiivseid parameetrite hindamismeetodeid nagu näiteks vähimruutude meetod.

3.5. Algoritm Google otsingu märksõnade sageduse andmete kaasamiseks prognoosimudelitesse.

Käesoleva töö käigus on koostatud praktiline ning stabiilne tegevusjuhend ehk algoritm, kuidas kaasata Google otsingu märksõnade sageduse andmeid prognoosimise teostamiseks nii, et see vähendaks prognoosimise viga (joonis 32):



Joonis 32. Tegevusjuhend ehk algoritm Google otsingu märksõnade sageduse andmete kaasamiseks prognoosimise teostamiseks

Antud juhendi põhjal on võimalik üles ehitada (vabavara) tarkvara rakendus, mis reaajas tegeleks majandusnäitajate prognoosimisega, laadides alla lähteandmestiku ehk Google otsingu märksõnade sageduse andmed. Sellisel viisil võib see lihtsusta oluliselt otsustamisprotsesse näiteks ettevõtte või ka riigiasutuste tasandil. Eestis autorile teadaolevalt taolisi automaatseid reaajasüsteeme veel ei ole, ehkki mujal maailmas taoliste tarkvara rakendustega tegeletud (Edmondson, Berhane).

KOKKUVÕTE

Internet on globaalne. Google otsingumootor on Internetis kõige levinum otsingumootor, mille abil sooritatakse miljardeid päringuid iga päev. Päringutes esinenud otsingu märksõnade sageduse üle peetakse arvestust. Kogutud andmed on kättesaadavad kõigile soovijatele. Kuna Interneti-otsingute andmete kasutamise kohta on teostatud märkimisväärne hulk varasemaid töid ja näib, et trend, kuidas kasutada neid andmeid mõtestatult jätkub, siis on tekkinud huvi testida Google otsingute märksõnade sageduse andmete kasutatavust ka Eesti jaoks. Käesolevas töös uuriti, kas Internetis teostatud otsingu märksõnade sageduse andmetest on abi eesti majandusprotsesside prognoosimisel.

Töö esimeses peatükis käsitleti teemaga seotud valdkondi ja sihtrühmi ning anti ülevaade varasemalt teostatud töödest. Teises peatükis esitati nelja vaatluse all oleva näitaja statistilised andmed. Seejärel selgitati Google Trends rakenduse toimimist, kirjeldati vaatlustulemuse kogumist ja moodustati Google indeks (GI) iga vaatluse all oleva näitaja jaoks. Anti ülevaade võimalikest prognoosimise lähenemistest. Kolmandas peatükis teostati Google indeksi ja majandusprotsesside näitajate graafilised analüüsid, mida kasutati otsingu märksõnade komplekti moodustamisel. Lisaks selgitati prognoosimodelite ülesehitust ja koostati prognoosimodelid. Seejärel esitleti mineviku (*ex post*) prognoosimise arvutustulemusi, tehti järeldused ja pakuti välja tuleviku (*ex ante*) prognoosid. Lõpetuseks formuleeriti algoritm Google otsingute andmete kaasamiseks prognoosimodelitesse. Varasemate töödega võrreldes tasub välja toomist käesolevas töös Google andmete ebastabiilsuse elimineerimist ja prognoosimodelite tihendamist ehk prognoosimodelite komponentide lisamist ning horisondi pikendamist ehk mitu-sammu-ette prognoosimise sammude arvu suurendamist. Prognoosimodelite parameetrite hinnangud leiti prognoosivea minimeerimise meetodil (*prediction error method* PEM).

Töös vaadeldi majanduse valdkondi, mille vastu eksisteerib laiema elanikkonna kõrgendatud huvi ehk valdkonnaga seotud märksõnade kohta on Google Trendi rakenduses analüüsi teostamiseks piisavalt andmeid. Töös selgus, et Google otsingu märksõnade sageduse andmete kaasamine prognoosimise teostamisel sõltub konkreetsest vaatluse all olevast majandusnäitajast. Nii töötuse määra baasmudeli kui ka laiendatud mudeli poolt prognoositud ajahetkede arväärtuste vead võrreldes tegeliku töötuse määra arväärtustega jäid suurusjärgus 10% piiridesse. Samas oli baasmudeli maksimaalne viga 12,82% ja minimaalne 0,58%. Laiendatud mudeli vead hälbisid aga 0,71% ja 9,62% vahemikus. Baasmudeli absoluutne keskmine viga oli 6,71% ja laiendatud mudelil vastavalt 5,18%. Sellest tulenevalt oli ka töötuse määra laiendatud mudeli keskmine absoluutne suhteline viga 1,53% võrra väikesem baasmudeli omast. Majutatute arvu laiendatud mudel ei vähendanud keskmiselt prognoosimise viga baasmudeliga võrreldes. Põhjus võis leiduda selles, et baasmudeli keskmine absoluutne suhteline viga 2,87% võrreldes tegeliku majutatud arvuga oli juba piisavalt madal ja laiendatud mudeli keskmine absoluutne suhteline viga 3,11% jäi sellele 0,24% võrra alla. Jaemüügi posti või Interneti teel puhul suutis laiendatud mudel järgida tegelike andmete trendi, baasmudel aga mitte. Iseseisvalt ei olnud küll jaemüügi posti või Interneti teel laiendatud mudel võrreldes teiste töös kasutatavate näitajatega kuigi täpne: keskmine absoluutne suhteline viga võrreldes tegelike andmetega oli 7,08%. Kuna aga baasmudeli keskmine absoluutne suhteline viga oli 12,23%, siis kokkuvõtvalt suutis laiendatud mudel edastada baasmudelit keskmiselt 5,15% võrra. Nii kinnisvara ostu-müügitehingute arvu baasmudel kui ka laiendatud mudel andsid sarnaseid tulemusi, suutes oma prognoosides järgida tegeliku kinnisvara ostu-müügilepingute arvu muutuste trendi. Laiendatud mudel suutis kokkuvõttes edestada baasmudelit 3,37% võrra. Kinnisvara ostu-müügitehingute arvu baasmudeli absoluutne keskmine viga oli 7,79% ja laiendatud mudelil vastavalt 4,42%.

Üldjoontes on käesolevas töös leitud tulemused kooskõlas varasemate uuringutega. Töö tulemusi, mida oli varasemate töödega võimalik võrrelda, näitasid üldjuhul paremaid tulemusi võrreldes varasemates töödes esitatutega. Kokkuvõttes näeme, et kui töötuse määra, jaemüügi posti või Interneti teel ning kinnisvara ostu-müügitehingute arvu näitajatel on Google otsingu märksõnade sageduse andmete kaasamine prognoosimise teostamiseks põhjendatud, siis majutatute arvu näitajal ei ole Google otsingu märksõnade sageduse andmete kaasamine prognoosimise teostamiseks tõendatud. Kokkuvõtteks võib

väita, et Google otsingu märksõnade sageduse andmete kaasamine prognooside tegemisel üldjuhul omab võimekust prognoosimise vea vähendamiseks.

Teostatud tööst ja koostatud tegevusjuhendist võib olla abi Google otsingu märksõnade sageduse andmete kasutamise edasisel uurimisel ühe ja konkreetse majandusprotsessi prognoosimisel. Autor näeb jätkuvalt potentsiaali Eesti töötuse määra uurimisel, keskendudes just sobilike prognoosimudelite väljatöötamisele. Samuti on huvitav valdkond Google otsingu märksõnade sageduse andmete ja börsil kauplemise aktiivsuse vaheline võimalik seos. Töös saavutatud tulemustele toetudes võib soovitada Interneti otsingu märksõnade sageduse andmete kasutamist mõne muu majandusliku, sotsiaalse või ühiskondliku protsessi prognoosimiseks. Kindlasti oleks huvitavaks väljakutseks reaalaja tarkvara rakenduse loomine, mis tegeleks majandusnäitajate prognoosimisega ja toetuks töö tulemusena formuleeritud algoritmile. Loodav tööriist võib oluliselt lihtsustada otsustamisprotsesse näiteks ettevõtte või ka riigiasutuste tasandil. Kuigi antud töös ei kasutatud muid Interneti suurandmete allikaid, nagu näiteks sotsiaalmeedia keskkonnad (Twitter, Facebook vms.), võib varasemalt teostatud tööde najal julgustada näiteks poliitilise või sotsiaalsete protsesside prognoosimise uurimist, eelnimetatud keskkondade poolt pakutavate suurandmeid kasutades. Tehtud töö annab eelduse, et piisavalt suurt ühiskondliku kõlapinda omava valdkonna prognooside tegemisel võib Interneti suurandmetest kasu olla.

VIIDATUD ALLIKAD

1. Statistikaameti andmebaas. Majandus – Infotehnoloogia ja side – Infotehnoloogia leibkonnas. IT32: 16-74-aastased arvuti- ja Interneti-kasutajad isikute rühma järgi. [http://pub.stat.ee/px-web.2001/Dialog/varval.asp?ma=IT32&ti=16%2D74%2DAASTASED+ARVUTI%2D+JA+INTERNETIKASUTAJAD+ISIKUTE+R%DCHMA+J%C4RGI&path=../Database/Majandus/05Infotehnoloogia/04Infotehnoloogia_leibkonnas/&lang=2] 02.11.2016.
2. InternetLiveStats.com, [<http://www.internetlivestats.com>] 02.11.2016.
3. GO-Gulf, [<http://www.go-gulf.com/blog/online-time/>] 02.11.2016.
4. Statista Inc., [<https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/>] 02.11.2016.
5. Alexa Internet Inc., [<http://www.alexa.com/topsites/countries/EE>] 02.11.2016.
6. InternetLiveStats.com, 2016, [<http://www.internetlivestats.com/one-second/>] 02.11.2016.
7. Kalev Hannes Leetaru, 2017, [<http://kalevleetaru.com/>] 08.05.2017.
8. **Leetaru, K., H.**, Culturomics 2.0, 2011, [<http://firstmonday.org/article/view/3663/3040>] 16.05.2017.
[https://www.youtube.com/watch?v=Ccum_Cu9sP8] 08.05.2017.
9. GDELT Project, [<http://www.gdeltproject.org/about.html>] 08.05.2017.
10. Google Trends, [<https://trends.google.com/trends/explore?date=2007-01-01%202017-05-04&geo=US&q=Obama,McCain,Romney,Clinton,Trump>] 04.05.2017.
11. **Achrekar, H., Gandhe, A., Lazarus, R., Yu, S., Liu, B.**, Predicting Flu Trends using Twitter Data, 2011, CPNS 2011, [http://www.cs.uml.edu/~hachreka/SNEFT/images/CPNS_2011.pdf] 04.05.2017.

12. **Jahanbakhsh, K., Moon, Y.,** The Predictive Power of Social Media: On the Predictability of U.S. Presidential Elections using Twitter, 2014,
[<https://arxiv.org/pdf/1407.0622.pdf>] 04.05.2017.
13. **Bendler, J., Brandt, T., Wagner, S., Neumann, D.,** Investigating crime-to-twitter relationships in urban environments – facilitating a virtual neighborhood watch, ECIS 2014,
[<http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1142&context=ecis2014>] 04.05.2017.
14. **Zhang, X., Fuehres H., Gloor, P. A.,** Predicting Stock Market Indicators Through Twitter “I hope it is not as bad as I fear”, 2010, COINs2010,
[http://www.ickn.org/documents/COINs2010_Twitter4.pdf] 04.05.2017.
15. **Asur, S., Huberman, B. A.,** Predicting the Future With Social Media, 2010,
[<https://arxiv.org/pdf/1003.5699.pdf>] 04.05.2017.
16. **O’Connor, B., Balasubramanyan, R., Routledge, B. R., Smith, N. A.,** From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series, 2010
[<http://homes.cs.washington.edu/~nasmith/papers/oconnor+balasubramanyan+routledge+smith.icwsm10.pdf>] 04.05.2017.
17. **Collins, S., Sun, Y., Kosinski, M., Stillwell, D., Markuzon, N.,** Are You Satisfied with Life?: Predicting Satisfaction with Life from Facebook, 2015,
[http://web.cs.ucla.edu/~yzsun/papers/SBP15_Collins_Camera.pdf] 04.05.2017
18. **Ekström, A., Kurland, L., Farrokhnia, N., Castrén, M., Nordberg, M.,** Forecasting Emergency Department Visits Using Internet Data, 2014,
[https://www.researchgate.net/publication/269276374_Forecasting_Emergency_Department_Visits_Using_Internet_Data] 04.04.2017.
19. **Baldacci E.,** From Data to Knowledge, 2016,
[https://ec.europa.eu/eurostat/cros/content/presentations-7_en,
https://ec.europa.eu/eurostat/cros/system/files/from_data_to_knowledge_baldacci.pptx] 18.05.2017.
20. **Wirthmann, A.,** Data Revolution and Official Statistics, 2016,
[https://ec.europa.eu/eurostat/cros/content/presentations-7_en,
https://ec.europa.eu/eurostat/cros/system/files/data_revolution_and_official_statistics_v3_wirth.pptx] 18.05.2017.

21. **Kovachev B.**, Big Data and Macroeconomic Nowcasting, 2016,
[\[https://ec.europa.eu/eurostat/cros/content/presentations-7_en,](https://ec.europa.eu/eurostat/cros/content/presentations-7_en)
[https://ec.europa.eu/eurostat/cros/system/files/bigdatanowcasting_bogomil_.pptx\]](https://ec.europa.eu/eurostat/cros/system/files/bigdatanowcasting_bogomil_.pptx)
 18.05.2017.
22. **Golmajer M.**, Early estimates, 2016,
[\[https://ec.europa.eu/eurostat/cros/content/presentations-7_en,](https://ec.europa.eu/eurostat/cros/content/presentations-7_en)
[https://ec.europa.eu/eurostat/cros/system/files/early_estimates_golmajer_13-10-2016.pptx\]](https://ec.europa.eu/eurostat/cros/system/files/early_estimates_golmajer_13-10-2016.pptx) 18.05.2017.
23. **Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., Brilliant, L.**, Detecting influenza epidemics using search engine query data, Nature, 457(7232):1012–14., 2009,
[\[https://5harad.com/mse231/papers/ginsberg_et_al_flutrends.pdf\]](https://5harad.com/mse231/papers/ginsberg_et_al_flutrends.pdf) 31.10.2016.
24. **Choi, H., Varian, H. R.**, Predicting the Present with Google Trends, 2009a, Google Inc., 2009,
[\[https://static.googleusercontent.com/media/www.google.com/et//googleblogs/pdfs/google_predicting_the_present.pdf\]](https://static.googleusercontent.com/media/www.google.com/et//googleblogs/pdfs/google_predicting_the_present.pdf) 19.05.2017.
25. **Choi, H., Varian, H. R.**, Predicting Initial Claims for Unemployment Benefits, 2009b, Google Inc. 2009,
[\[http://static.googleusercontent.com/media/research.google.com/et//archive/papers/initialclaimsUS.pdf\]](http://static.googleusercontent.com/media/research.google.com/et//archive/papers/initialclaimsUS.pdf)] 31.10.2016
26. **Askatas, N., Zimmermann, K. F.**, Google Econometrics and Unemployment Forecasting, Applied Economics Quarterly, 55(2):107–120., 2009,
[\[http://ftp.iza.org/dp4201.pdf\]](http://ftp.iza.org/dp4201.pdf). 28.10.2016.
27. **Suhoy, T.**, Query Indices and a 2008 Downturn: Israeli Data, 2009, Bank of Israel Discussion Paper, 2009,
[\[https://www.boi.org.il/deptdata/mehkar/papers/dp0906e.pdf\]](https://www.boi.org.il/deptdata/mehkar/papers/dp0906e.pdf) 28.10.2016.
28. **D’Amuri, F.**, Predicting unemployment in short samples with Internet job search query data, MPRA Working Paper 18403, 2009
[\[https://mpra.ub.uni-muenchen.de/18403/1/MPRA_paper_18403.pdf\]](https://mpra.ub.uni-muenchen.de/18403/1/MPRA_paper_18403.pdf)
 31.10.2016.

29. **Guzman, G.**, Internet search behavior as an economic forecasting tool: The case of inflation expectations, *Journal of Economic and Social Measurement*, 36(2011), 119–167, 2011,
[<http://www.datascienceassn.org/sites/default/files/Internet%20search%20behavior%20as%20an%20economic%20forecasting%20tool%20the%20case%20of%20inflation%20expectations.pdf>] 31.10.2016.
30. **Preis, T., Moat, H. S., Stanley, H. E.**, Quantifying trading behavior in financial markets using Google Trends, *Scientific Reports*, 3. 1684. ISSN 2045-2322, 2013,
[<http://www.nature.com/articles/srep01684>] 01.11.2016.
31. Euroopa Komisjon, Horizon 2020, Big Data, 2016,
[https://ec.europa.eu/eurostat/cros/content/2016-2017-calls-big-data_en] 03.11.2016.
32. **Lindberg, F.**, Nowcasting Swedish Retail Sales with Google Search Query Data, Stockholm University, 2011.
33. **Tuhkuri, J.**, Big Data: Google Searches Predict Unemployment in Finland, *ETLA Reports* 31, 2014, [<https://www.etla.fi/wp-content/uploads/ETLA-Raportit-Reports-31.pdf>] 03.11.2016.
34. **Chmyznikov, G., Galvanauskas, L.**, Forecasting Economic Activity in the Baltics: Let us Google it, Stockholm School of Economics in Riga, Riga, 2013,
[http://www.sseriga.edu/download.php?file=files/researchPapers/chmyznikov_galvanauskas_1_2013.pdf] 02.11.2016.
35. **Tuhkuri, J.**, Forecasting unemployment with Google Searches, ETLA Working Papers No 35. 2016, [<https://www.etla.fi/wp-content/uploads/ETLA-Working-Papers-35.pdf>] 31.10.2016.
36. **Chadwick, M. G., Sengul, G.**, Nowcasting unemployment rate in Turkey: Let's ask Google, Central Bank of the Republic of Turkey Working Paper 12/18, (June), 2012, [<http://www.tcmb.gov.tr/wps/wcm/connect/ff056d6a-8781-426e-bb5c-0d90f61e8af4/WP1218.pdf?MOD=AJPERES&CACHEID=ROOTWORKSPACEff056d6a-8781-426e-bb5c-0d90f61e8af4>] 01.11.2016.
37. **Vosen, S., Schmidt, T.**, Forecasting private consumption: Survey-based indicators vs. Google Trends, *Journal of Forecasting* 30 (6): 565–578, 2009,

- [http://www.rwi-essen.de/media/content/pages/publikationen/ruhr-economic-papers/REP_09_155.pdf] 31.10.2016.
38. **Kulkarni, R.K. Haynes, R., Stough, and J.H.P. Paelinck.,** Forecasting housing prices with Google econometrics, mimeo, School of Public Policy, George Mason University, Fairfax, VA 22030, 2009, [https://www.researchgate.net/publication/228181453_Forecasting_Housing_Prices_with_Google_Econometrics] 31.10.2016.
 39. **Kholodilin, KA, Podstawski, M & Siliverstovs, B.,** Do Google searches help in nowcasting private consumption? A real-time evidence for the US., DIW DP 997, DIW, Berlin, 2010, [http://www.diw.de/documents/publikationen/73/diw_01.c.356220.de/dp997.pdf] 31.10.2016.
 40. **Moen, E., Anvik, C., Gjelstad K.,** Just Google It! Forecasting Norwegian unemployment figures with web queries. CREAM Publications;11/2010, 2010, [<http://hdl.handle.net/11250/95461>] 01.11.2016.
 41. **Bughin, J. R.,** Nowcasting' the Belgian Economy., SSRN working paper, 2011, [https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1903791] 01.11.2016.
 42. **McLaren, N., Shanbhogue, R.,** Using Internet search data as economic indicators, Bank of England Quarterly Bulletin, Q2:134–140, 2011, [<http://www.bankofengland.co.uk/publications/Documents/quarterlybulletin/qb110206.pdf>] 01.11.2016.
 43. **Pan, B., Wu, D. C., & Song, H.,** Forecasting Hotel Room Demand Using Search Engine Data, Journal of Hospitality and Tourism Technology, 3(3), 196-210, (2012), [<http://www.panb.people.cofc.edu/pan/BingPanDorisWuHaiyanSongForecastingHotelOccupancySearchEngineData.pdf>] 31.10.2016.
 44. **Wu, L., Brynjolfsson, E.,** The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales, In Goldfarb, A., Greenstein, S., and Tucker, C., editors, Economic Analysis of the Digital Economy, pages 89–118. University of Chicago Press, 2014, [<http://www.nber.org/chapters/c12994.pdf>] 31.10.2016.

45. **Pavlicek J, Kristoufek L.**, Nowcasting Unemployment Rates with Google Searches: Evidence from the Visegrad Group Countries, PLoS ONE 10(5): e0127084, 2015,
[<http://journals.plos.org/plosone/article/asset?id=10.1371/journal.pone.0127084.PDF>] 01.11.2016.
46. Eurostat [<http://ec.europa.eu/eurostat/web/lfs/data/database>] 28.01.2017.
47. Statistikaamet, Majutatud - Näitaja, Aasta ning Kuu. Turism, majutus ja toitlustus - Majutus. TU121: Majutatud (kuud). [http://pub.stat.ee/px-web.2001/Dialog/varval.asp?ma=TU121&ti=MAJUTATUD+%28KUUD%29&path=../Database/Majandus/23Turism_ja_majutus/02Majutus/&lang=2] 28.01.2017.
48. Statistikaamet, Kaupade jaemüük tegevusala (EMTAK 2008) järgi (kvartalid). Majandus – Sisekaubandus. KM0101: kaupade jaemüük tegevusala (EMTAK 2008) järgi (kvartalid). [http://pub.stat.ee/px-web.2001/Dialog/varval.asp?ma=KM0101&ti=KAUPADE+JAEM%DC%DCK+TEGEVUSALA+%28EMTAK+2008%29+J%C4RGI+%28KVARTALID%29&path=../Database/Majandus/17Sisekaubandus/04Sisekaubanduse_majandusnaitajad/&lang=2] 28.01.2017.
49. Statistikaamet, Notariaalselt tõestatud ostu-müügilepingud liigi järgi (kvartalid) Majandus – Kinnisvara - Kinnisvaratehingud. KV015: Notariaalselt tõestatud ostu-müügilepingud liigi järgi (kvartalid). [<http://pub.stat.ee/px-web.2001/Dialog/varval.asp?ma=KV015&ti=NOTARIAALSELT+T%D5ESTATUD+OSTU%2DM%DC%DCGILEPINGUD+LIIGI+J%C4RGI+%28KVARTALID%29&path=../Database/Majandus/09Kinnisvara/04Kinnisvaratehingud/&lang=2>] 28.01.2017.
50. npm, Inc., [<https://www.npmjs.com/package/google-trends-api>] 19.05.2017.
51. Google Trends, [<https://trends.google.com/trends/explore?date=2004-01-01%202016-12-31&q=t%C3%B6%C3%B6>] 11.03.2017.
52. Google Trends, [<https://trends.google.com/trends/explore?date=2004-01-01%202016-12-31&q=majutus>] 11.03.2017.
53. Google Trends, [<https://trends.google.com/trends/explore?date=2004-01-01%202016-12-31&q=e-pood>] 11.03.2017.

54. Google Trends, [<https://trends.google.com/trends/explore?date=2004-01-01%202016-12-31&q=city24>] 11.03.2017.
55. Yaffee, R. A., Forecast evaluation with Stata, London, 2010, [<http://www.stata.com/meeting/uk10/UKSUG10.Yaffee.pdf>] 31.03.2017.
56. Kangro R., E-kursuse "Aegridade Analüüs" materjalid, Tartu, 2012, [http://dspace.ut.ee/bitstream/handle/10062/27703/aegridede_analuus.pdf] 02.04.2017.
57. Stoica, P., Nehorai, A., On multistep prediction error methods for time series models, Journal of Forecasting, Volume 8, Issue 4, October/December 1989, pp 357–368 [<http://onlinelibrary.wiley.com/doi/10.1002/for.3980080402/abstract>] [<http://web.a.ebscohost.com.ezproxy.utlib.ut.ee/ehost/pdfviewer/pdfviewer?sid=aab33d90-f55a-4dcd-9551-69a653a542b9%40sessionmgr4006&vid=3&hid=4204>] 18.05.2017.
58. Pelckmans, K., System Identification, Uppsala University, Information Technology, Lecture Notes, 2012, [<https://www.it.uu.se/edu/course/homepage/systemid/vt12>] 19.05.2017. [<https://www.it.uu.se/edu/course/homepage/systemid/vt12/ch6.pdf>] 19.05.2017.
59. Edmondson, M., My Google Analytics Time Series Shiny App (Alpha), 2014, [<http://markedmondson.me/my-google-analytics-time-series-shiny-app-alpha>] 18.05.2017.
60. Berhane, F., Google Trends Analytics using Shiny [http://datascience-enthusiast.com/R/gtrends_shiny.html] 18.05.2017.

LISAD

Lisa 1. Statistika metoodikad

1. Töötuse määr

Põhimõisted

Majanduslikult aktiivne rahvastik ehk tööjõud – isikud, kes soovivad töötada ja on võimelised töötama (hõivatute ja töötute summa).

Majanduslikult passiivne ehk mitteaktiivne rahvastik – isikud, kes ei soovi töötada või ei ole selleks võimelised.

Tööealine rahvastik – rahvastiku majandusliku aktiivsuse uurimisel aluseks võetavas ehk tööjõu-uuringu objektiks olevas vanusevahemikus rahvastik (15–74-aastased).

(Tööga) hõivatu – isik, kes uuritaval perioodil:

- töötas ja sai selle eest tasu kas palgatöötajana, ettevõtjana või vabakutselisena,
- töötas otsese tasuta pereettevõttes või oma talus,
- ajutiselt ei töötanud.

Tööhõive määr – hõivatute osatähtsus tööealises rahvastikus.

Tööjõus osalemise määr (aktiivsuse määr) – tööjõu osatähtsus tööealises rahvastikus.

Töötu – isik, kelle puhul on korraga täidetud kolm tingimust:

- on ilma tööta (ei tööta mitte kusagil ega puudu ajutiselt töölt),
- on töö leidmisel valmis kohe (kahe nädala jooksul) tööd alustama,
- otsib aktiivselt tööd.

Töötuse määr ehk tööpuuduse määr – töötute osatähtsus tööjõus.

Metoodika

Statistikaamet korraldas esimese Eesti tööjõu-uuringu küsitluse 1995. aasta alguses (ETU 95). 1997.–1999. aastal toimus küsitlus II kvartalis. Alates 2000. aastast on tööjõu-uuring kvartali- ja aastaväljundiga pidevuuring. Valimisse sattunud leibkondade uuringunädalad on jaotatud ühtlaselt üle terve aasta, s.t aasta iga nädala kohta korraldatakse võrdne arv küsitlusi.

Kuni 2000. aasta II kvartalini koosnes ETU küsimustik kahest osast:

- uuringunädala osa, milles koguti tööalast infot küsitlusele eelnenud nädala kohta,
- tagasivaateline osa, milles koguti andmeid töömuutuste kohta uuringule eelnenud aastatel.

Kvartalitulemusi saadakse pidevuuringu põhjal alates 2000. aasta I kvartalist, aga sarnaselt eelmiste küsitlustega koguti I ja II kvartalis lisaks tagasivaatelisi andmeid 1999. aasta kohta, sest siis toimus küsitlus vaid II kvartalis. Alates 2000. aasta III kvartalist tagasivaatelisi andmeid enam ei koguta. Selle järele puudub vajadus, sest küsitlus toimub pidevalt ja kõik andmed saadakse uuringunädala järgi. Aastakeskmised on kvartalitulemuste aritmeetilised keskmised.

Küsitletavate valiku aluseks oli 1995. aastal 1989. aasta rahvaloenduse andmebaas, 1997.–2001. aastal AS Andmevara rahvastiku andmebaas (rahvastikuregister), 2002.–2004. aastal 2000. aasta rahva ja eluruumide loenduse aadressbaas, 2005.–2012. aastal rahvastikuregister, alates 2013. aastast 2011. aasta rahva ja eluruumide loenduse aadressbaas.

Tööjõu-uuringuga kogutakse andmeid tööelistelt isikutelt (15–74-aastased). Enne 2002. aastat arvestati küsitletava vanuse määramisel vaid sünniaastat: tööelised on inimesed, kes uuringuaasta 1. jaanuaril olid 15–74-aastased. 2002. aastast alates arvestatakse ka sünnikuupäeva: tööelised on inimesed, kes uuringunädalal olid 15–74-aastased.

Sünnikuupäeva arvestamine vanuse arvutamisel alates 2002. aastast võimaldab täpselt arvutada ka vanuserühma “16 kuni pensioniiga” andmeid. Enne 2002. aastat arvestati

töøjõu-uuringus pensioniiga ainult täisaastates. Töøjõu-uuringus kasutatud pensioniiga on esitatud järgmises tabelis.

Tabel 1. Töøjõu-uuringus kasutatud pensioniiga

	Vanus, aastat	
	mehed	naised
1989–1994	60	55
1995–1996	61	56
1997–1998	62	57
1999–2001	63	58
2002–2003	63	58,5
2004	63	59
2005–2006	63	59,5
2007	63	60
2008–2009	63	60,5
2010	63	61
2011–2012	63	61,5
2013	63	62
2014–2015	63	62,5

Allikas: (Statistikaamet 2016)

Et ETU 95 ja ETU 97 uuringunädala andmed on isikute kohta, kes uuringuaasta 1. jaanuaril olid 15–74-aastased, ja tagasivaatelse osa andmed on kogutud täpselt samadelt isikutelt, kes olid aga varasematel aastatel teistes vanusevahemikes, siis tagamaks andmete võrreldavust aastate kaupa ning tavapäraselt rahvastikuandmete esitust viieaastaste vanuserühmadena, on 1989.–1996. aasta kohta kogutud tagasivaatelised andmed esitatud 15–69-aastaste kohta.

Alates 1998. aastast kuni 2000. aasta II kvartalini küsitleti töøjõu-uuringus ka 75-aastasi inimesi. Nii saadi 1997.–1999. aasta kohta tagasivaatelised andmed, mis hõlmavad 15–74-aastasi. Kõik üle 74-aastased inimesed on 1997. ja järgmiste aastate andmetes loetud mitteaktiivseks, varasemate aastate puhul üle 69-aastased.

Töøjõu-uuringu andmete laiendamisel on aluseks arvestuslik rahvaarv uuringuaasta 1. jaanuaril. Laiendustegurid arvutatakse maakonna, soo ja 5-aastaste vanuserühmade järgi. Andmebaasis avaldatud 1990.–1999. aasta andmed on täpsustatud, kasutades 1989. ja 2000. aasta rahvaloenduse vaheliste aastate korrigeeritud rahvaarve. 2000.–2013. aasta

andmed on täpsustatud, kasutades 2011. aasta rahvaloenduste ja registrite andmete tuginedes ümberarvutatud rahvaarve (vt rahvaarvu ümberarvutuste metoodikast <http://www.stat.ee/75541> ja tööjõu-uuringu ümberarvutuste metoodikast <http://www.stat.ee/76254>).

Valimi laiendamisel üldkogumi kohta saadud andmed on kõik tegelike parameetrite hinnangud. Hinnanguid, mis põhinevad valimi vähem kui 20 isikul, ei ole avaldatud (tabelites tähistus “..”), sest need ei ole usaldusväärsed.

Ümardamise tõttu ei võrdu tabelites arvude liitmise tulemus alati summaarse näitavuga. Erinevus võib olla kuni paar viimase koha ühikut.

Alates 2015. aasta I kvartalist ei avalda Statistikaamet enam tööjõu-uuringu tulemusi tegevus- ja ametialade järgi kvartali kaupa, vaid avaldab need korra aastas aasta kokkuvõttena. Kvartaalsed tööhõiveandmed tegevusalade järgi on ka edaspidi kättesaadavad Statistikaameti avaldatavas ettevõtlusstatistikas, vabade ja hõivatud ametikohtade ning rahvamajanduse arvepidamise statistikas.

(Allikas: Statistikaamet 2016)

Statistikaameti andmebaas. Sotsiaalelu – Tööturg. [http://pub.stat.ee/px-web.2001/Database/Sotsiaalelu/15Tooturg/12Tooturu_uldandmed/04Luhiajastatistika/T_T_047.htm] 15.09.2016.

2. Majutatud

Põhimõisted

Majutatu – turist ehk ööbiv külastaja, kes veedab vähemalt ühe öö majutusettevõttes.

Majutusettevõtte – majandusüksus, mille kaudu osutab ettevõtja oma majandus- või kutsetegevusega majutusteenust. Majutusettevõtte liigid on järgmised: hotell, motell, külalistemaja, hostel, puhkeküla ja -laager, puhkemaja, külaliskorter ning kodumajutus. Turistide majutusüksused on jaotatud kahte põhirühma: kollektiivmajutus (teenindavad turiste äriüksustena) ja eramajutus. Majutusstatistika kajastab ainult kollektiivmajutuskohtade andmeid.

Metoodika

Majutusettevõtete üldkogum hõlmab ettevõtteid, mille põhi- või kõrvaltegevusala Eesti majanduse tegevusalade klassifikaatori (EMTAK 2008) järgi on:

- 551 – hotellid ja muu sarnane majutus,
- 552 – puhkuse- ja muu lühiajaline majutus,
- 553 – laagriplatsid, vagunelamute ja haagissuvilate parkimisplatsid,

ning Majandustegevuse registrisse (MTR) kantud majutusettevõtteid. Vaadeldakse vähemalt viie voodikohaga majutusettevõtteid. Alates 2002. aastast hõlmab majutusstatistika ka majutusteenust pakkuvaid taastusravikeskusi.

(Allikas: Statistikaamet 2016)

Statistikaameti andmebaas. Majandus – Turism, majutus ja toitlustus - Majutus.

[http://pub.stat.ee/px-web.2001/Database/Majandus/23Turism_ja_majutus/02Majutus/TU_11.htm]

15.09.2016.

3. Jaemüük posti või Interneti teel

Põhimõisted

Ettevõtte tegevusala – ettevõtte põhitegevusala Eesti majanduse tegevusalade klassifikaatori (EMTAK) järgi.

Jaekaubandusettevõtte – ettevõtte, mille põhitegevus on eelkõige kauba müümine (v.a mootorsõidukid ja mootorrattad) peamiselt kodumajapidamises kasutamiseks (Eesti majanduse tegevusalade klassifikaatoris (EMTAK) tegevusala G 47).

Müügitulu – nii põhi- kui ka kõrvaltegevusena valmistatud toodete, osutatud teenuste ja edasimüügi eesmärgil soetatud kaupade müügist saadud või saadaolev tulu, mis ei sisalda käibemaksu ega aktsiise. Müügitulu võrdub saadud või saadaoleva tasuga ja arvestatakse tekkepõhiselt. Müügitulu on omandiõiguse üleminek ostjale. Müügitulusse arvestatakse ka taara ja pakkematerjali maksumus, samuti komisjoni- ja agendilepingute teenustasud (komisjonitasud). Müügitulu ei sisalda põhivara müügitulu; muud äritulu (põhivara likvideerimise tulu, trahvi- ja viivisetulu, kasum ostjate ja tarnijate valuutakursi

muutustest ja aegunud kreditoorsest võlgnevusest, aruandeperioodil selgunud eelmiste aastate tulu, saadud annetused jms); finants- ega erakorralist tulu; riigi- või kohalikust eelarvest või Euroopa Liidu institutsioonidelt saadud dotatsioone (tegevuskulude sihtfinantseerimine); käibemaksukohustuslasel käibemaksu; tootjaettevõttel ja aktsiisilaopidajal aktsiise; teiste nimel sissenõutud summasid (komisjoni- ja agendilepingud).

Jaemüük posti või Interneti teel (G 4791) - Toodete jaemüük posti või Interneti teel, st müük, mille puhul ostja teeb oma valiku reklaamide, kataloogide, veebilehel pakutava informatsiooni ja kaubanäidiste vms põhjal ning esitab ostutellimuse posti või telefoni teel või Interneti kaudu (tavaliselt läbi spetsiaalse veebilehe). Ostetud tooted saab veebilehelt alla laadida või saadetakse need tarbijale posti teel:

- mis tahes toodete jaemüük posti teel,
- mis tahes toodete jaemüük Interneti teel,
- otsemüük televisiooni, raadio või telefoni teel,
- jaemüük Interneti-oksjonite kaudu.

Siia ei kuulu:

- mootorsõidukite, nende osade ja lisaseadmete jaemüük Interneti teel, vt grupid 451, 453,
- mootorrataste, nende osade ja lisaseadmete jaemüük Interneti teel, vt grupp 45401 (Registrite ja Infosüsteemide Keskus 2016).

Metoodika

Andmed kogutakse statistikatöö 20007 „Ettevõtete majandusnäitajad“ küsimustikuga „Majandustegevus“. Aruandeperiood on kvartal.

Üldkogumi moodustavad ettevõtted, mis on kantud Justiitsministeeriumi Registrikeskuse andmebaasi andmetel moodustatud majanduslikult aktiivsete ettevõtete nimekirja ehk statistilisse profiili.

Töoga hõivatud isikute arvu põhjal kuuluvad vaatluse alla kõik vähemalt 20 hõivatuga ettevõtted. Väikeettevõtete (1–19 tööga hõivatut) andmed kogutakse valikuuringu

meetodil, kasutades stratifitseeritud lihtsat juhuslikku valikut. Üldkogum stratifitseeritakse ettevõtte tegevusala ja tööga hõivatute arvu järgi.

Valikuuringuga kogutud andmete põhjal hinnatakse üldkogumi näitajaid. Küsimustiku esitamata jätnud ettevõttele asendatakse võimaluse korral varasema perioodi andmed. Kui küsimustiku esitamata jätnud ettevõtte kohta andmed puuduvad, asendatakse tema andmed tegevusala grupi vastavate andmete keskmistega.

(Allikas: Statistikaamet 2016, Registrite ja Infosüsteemide Keskus 2016)

Statistikaameti andmebaas. Majandus – Sisekaubandus. [http://pub.stat.ee/px-web.2001/Database/Majandus/17Sisekaubandus/04Sisekaubanduse_majandusnaitajad/KM_0034.htm] 15.09.2016.

Registrite ja Infosüsteemide Keskus. e-äriregister. EMTAK tegevusalad [http://www.rik.ee/sites/www.rik.ee/files/elfinder/article_files/EMTAK%202008%20EST%20-%20selgitavate%20m%C3%A4rkustega.pdf] 13.10.2016.

4. Kinnisvara ostu-müügitehingute arv.

Põhimõisted

Kinnisasi – maatükk koos selle oluliste osadega. Maatüki olulised osad on sellega püsivalt ühendatud asjad (ehitised, kasvav mets, koristamata vili jne) ja maatükiga seotud asjaõigused.

Kinnisvara – isikule kuuluvad kinnisasjad ja kinnisasjadega seotud rahaliselt hinnatavad õigused ning kohustused, sh hoonestusõigus ja kasutusvaldus.

Kinnisvara ostu-müügitehing – ostu-müügitehing, mille objekt on kinnistusraamatusse kantud. Siia kuuluvad:

- hoonestamata kinnistu ostu-müügitehingud, mille objekt on ainult maatükk,
- hoonestatud kinnistu ostu-müügitehingud, mille objekt on maatükk koos olemasolevate või ehitusjärgus hoonetega.

Metoodika

Andmed kinnisvara ostu-müügitehingute arvu ja väärtuse kohta kogutakse kõikse vaatlusega "Kinnisvara". Alates 2008. aastast on kinnisvara ostu-müügitehingute riikliku statistika (tabelid KV015, KV02, KV042, KV05) tootmise alus Maa-ameti avaliku andmebaasi andmed. Andmete avaldamisel lähtutakse tehingute andmebaasis kasutatavast konfidentsiaalsuse reeglist: tehingute väärtus kuvatakse vaid juhul, kui on toimunud vähemalt 5 tehingut. Varasematel aastatel kogus Statistikaamet andmeid notaritelt.

(Allikas: Statistikaamet 2016)

Statistikaameti andmebaas. Majandus – Kinnisvara - Kinnisvaratehingud
[http://pub.stat.ee/px-web.2001/Database/Majandus/09Kinnisvara/04Kinnisvaratehingud/KV_05.htm]
15.09.2016.

Lisa 2. Prognoosmodelite valemid

Kvartali detailsusega baasmudelid

Prognoositud aegridade väärtused $y_{p\ t}$ arvutatakse kasutades mudeleid:

$$M(0.1) : \log(y_{p\ t}) = \beta_{01} + \beta_{11} \log(y_{t-1}) + \beta_{21} \log(y_{t-2}) + \cdots + \beta_{41} \log(y_{t-4})$$

$$M(0.2) : \log(y_{p\ t}) = \beta_{02} + \beta_{12} \log(y_{t-2}) + \beta_{22} \log(y_{t-3}) + \cdots + \beta_{42} \log(y_{t-5})$$

...

$$M(0.5) : \log(y_{p\ t}) = \beta_{05} + \beta_{15} \log(y_{t-5}) + \beta_{25} \log(y_{t-6}) + \cdots + \beta_{45} \log(y_{t-8})$$

ehk:

$$M(0.h) : \log(y_{p\ t}) = \beta_{0h} + \beta_{1h} \log(y_{t-h}) + \beta_{2h} \log(y_{t-(h+1)}) + \cdots + \beta_{4h} \log(y_{t-(h+3)})$$

ehk:

$$(10) \quad M(0.h) : \log(y_{p\ t}) = \beta_{0h} + \sum_{i=1}^4 \beta_{ih} \log(y_{t-(h+(i-1))})$$

Kvartali detailsusega laiendatud mudelid

Prognoositud aegridade väärtused $y_{p\ t}$ arvutatakse kasutades mudeleid:

$$M(1.1) : \log(y_{p\ t}) = \alpha_{01} + \alpha_{11} \log(y_{t-1}) + \alpha_{21} \log(y_{t-2}) + \cdots + \alpha_{41} \log(y_{t-4}) + \\ + \gamma_{01} \log(x_t) + \gamma_{11} \log(x_{t-1}) + \gamma_{21} \log(x_{t-2}) + \cdots + \gamma_{41} \log(x_{t-4})$$

$$M(1.2) : \log(y_{p\ t}) = \alpha_{02} + \alpha_{12} \log(y_{t-2}) + \alpha_{22} \log(y_{t-3}) + \cdots + \alpha_{42} \log(y_{t-5}) + \\ + \gamma_{02} \log(x_{t-1}) + \gamma_{12} \log(x_{t-2}) + \gamma_{22} \log(x_{t-3}) + \cdots + \gamma_{42} \log(x_{t-5})$$

...

$$M(1.5) : \log(y_{p\ t}) = \alpha_{05} + \alpha_{15} \log(y_{t-5}) + \alpha_{25} \log(y_{t-6}) + \cdots + \alpha_{45} \log(y_{t-8}) + \\ + \gamma_{05} \log(x_{t-4}) + \gamma_{15} \log(x_{t-5}) + \gamma_{25} \log(x_{t-6}) + \cdots + \gamma_{45} \log(x_{t-8})$$

ehk:

$$M(1.h) : \log(y_{p\ t}) = \alpha_{0h} + \alpha_{1h} \log(y_{t-h}) + \alpha_{2h} \log(y_{t-(h+1)}) + \cdots + \alpha_{4h} \log(y_{t-(h+3)}) + \\ + \gamma_{0h} \log(x_{t-(h-1)}) + \gamma_{1h} \log(x_{t-h}) + \gamma_{2h} \log(x_{t-(h+1)}) + \cdots + \gamma_{4h} \log(x_{t-(h+3)})$$

ehk:

$$(11) \quad M(1.h) : \log(y_{p\ t}) = \alpha_{0h} + \sum_{i=1}^4 \alpha_{ih} \log(y_{t-(h+(i-1))}) + \sum_{j=0}^4 \gamma_{jh} \log(x_{t-(h+(j-1))})$$

Kuu detailsusega baasmudelid

Prognoositud aegridade väärtused $y_{p\ t}$ arvutatakse kasutades mudeleid:

$$M(0.1) : \log(y_{p\ t}) = \beta_{01} + \beta_{11} \log(y_{t-1}) + \beta_{21} \log(y_{t-2}) + \cdots + \beta_{121} \log(y_{t-12})$$

$$M(0.2) : \log(y_{p\ t}) = \beta_{02} + \beta_{12} \log(y_{t-2}) + \beta_{22} \log(y_{t-3}) + \cdots + \beta_{122} \log(y_{t-13})$$

...

$$M(0.15) : \log(y_{p\ t}) = \beta_{015} + \beta_{115} \log(y_{t-15}) + \beta_{215} \log(y_{t-16}) + \cdots + \beta_{1215} \log(y_{t-26})$$

ehk:

$$M(0.h) : \log(y_{p\ t}) = \beta_{0h} + \beta_{1h} \log(y_{t-h}) + \beta_{2h} \log(y_{t-(h+1)}) + \cdots + \beta_{12h} \log(y_{t-(h+11)})$$

ehk:

$$(12) \quad M(0.h) : \log(y_{p\ t}) = \beta_{0h} + \sum_{i=1}^{12} \beta_{ih} \log(y_{t-(h+(i-1))})$$

Kuu detailsusega laiendatud mudelid

Prognoositud aegridade väärtused $y_{p\ t}$ arvutatakse kasutades mudeleid:

$$\begin{aligned} \text{M(1.1)} : \quad & \log(y_{p\ t}) = \alpha_{01} + \alpha_{11} \log(y_{t-1}) + \alpha_{21} \log(y_{t-2}) + \cdots + \alpha_{121} \log(y_{t-12}) + \\ & + \gamma_{-11} \log(x_{t+1}) + \gamma_{01} \log(x_t) + \gamma_{11} \log(x_{t-1}) + \gamma_{21} \log(x_{t-2}) + \cdots + \gamma_{121} \log(x_{t-12}) \end{aligned}$$

$$\begin{aligned} \text{M(1.2)} : \quad & \log(y_{p\ t}) = \alpha_{02} + \alpha_{12} \log(y_{t-2}) + \alpha_{22} \log(y_{t-3}) + \cdots + \alpha_{122} \log(y_{t-13}) + \\ & + \gamma_{-12} \log(x_t) + \gamma_{02} \log(x_{t-1}) + \gamma_{12} \log(x_{t-2}) + \gamma_{22} \log(x_{t-3}) + \cdots + \gamma_{122} \log(x_{t-13}) \end{aligned}$$

...

$$\begin{aligned} \text{M(1.15)} : \quad & \log(y_{p\ t}) = \alpha_{015} + \alpha_{115} \log(y_{t-15}) + \alpha_{215} \log(y_{t-16}) + \cdots + \alpha_{1215} \log(y_{t-26}) + \\ & + \gamma_{-115} \log(x_{t-13}) + \gamma_{015} \log(x_{t-14}) + \gamma_{115} \log(x_{t-15}) + \gamma_{215} \log(x_{t-16}) + \cdots + \gamma_{1215} \log(x_{t-26}) \end{aligned}$$

ehk:

$$\begin{aligned} \text{M(1.h)} : \quad & \log(y_{p\ t}) = \alpha_{0h} + \alpha_{1h} \log(y_{t-h}) + \alpha_{2h} \log(y_{t-(h+1)}) + \cdots + \alpha_{12h} \log(y_{t-(h+11)}) + \\ & + \gamma_{-1h} \log(x_{t-(h-2)}) + \gamma_{0h} \log(x_{t-(h-1)}) + \gamma_{1h} \log(x_{t-h}) + \gamma_{2h} \log(x_{t-(h+1)}) + \cdots + \gamma_{12h} \log(x_{t-(h+11)}) \end{aligned}$$

ehk:

$$(13) \quad \text{M(1.h)} : \log(y_{p\ t}) = \alpha_{0h} + \sum_{i=1}^{12} \alpha_{ih} \log(y_{t-(h+(i-1))}) + \sum_{j=-1}^{12} \gamma_{jh} \log(x_{t-(h+(j-1))})$$

kus $y_{p\ t}$ – prognoositud aegrea arvuline väärtus ajahetkel t ,
 y_t – tegeliku aegrea väärtus ajahetkel t ,
 x_t – Google indeksi väärtus ajahetkel t ,
 $\beta_{0h}; \beta_{ih}$ – baasmudeli parameetrid,
 $\alpha_{0h}; \alpha_{ih}$ – laiendatud mudeli (tegeliku aegrea osa) parameetrid,
 γ_{jh} – laiendatud mudeli (Google indeksi aegrea osa) parameetrid,
 h – prognoositava ajaperioodi järk ehk horisont,
 $i; j$ – parameetrite indeksid.

Lisa 3. Joonas Tuhkuri tööde prognoosid

Tuhkuri, J. (2014)

Tabel 1. Soome töötuse määra prognooside vead

	Malli	MAE	Δ
t	(0)	7.8 %	10.0 %
	(1)	7.0 %	
$t + 1$	(0)	9.3 %	16.9 %
	(1)	7.7 %	
$t + 2$	(0)	10.5 %	32.9 %
	(1)	7.0 %	
$t + 3$	(0)	11.1 %	39.2 %
	(1)	6.7 %	
$t + 4$	(0)	11.3 %	30.5 %
	(1)	7.7 %	
$t + 5$	(0)	11.3 %	25.3 %
	(1)	8.4 %	
$t + 6$	(0)	11.4 %	20.5 %
	(1)	9.0 %	

MAE = absoluuttinen keskivirhe

Δ = ennustetarkkuuden muutos

Taulukko 4: Mallin (0) ja (1) nykyhetken ja tulevaisuuden ennustetarkkuus

Allikas: (Tuhkuri, J., Big Data: Google Searches Predict Unemployment in Finland, ETLA Reports 31, 2014, [<https://www.etla.fi/wp-content/uploads/ETLA-Raportit-Reports-31.pdf>] 03.11.2016).

Parendus on välja toodud muutuse protsendina.

Tuhkuri, J. (2016)

Tabel 2. Ameerika Ühendriikide töötuse määra prognooside vead

Horizon	Model	MAPE	Δ
$h = 0$	(0.0)	4.58%	4.32%
	(1.0)	4.38%	
$h = 1$	(0.0)	7.57%	7.48%
	(1.1)	7.01%	
$h = 2$	(0.0)	9.48%	-3.92%
	(1.2)	9.85%	
$h = 3$	(0.0)	10.4%	-6.28%
	(1.3)	11.06%	
$h = 4$	(0.0)	11.1%	-17.22%
	(1.4)	13.02%	
$h = 5$	(0.0)	11.96%	-13.22%
	(1.5)	13.54%	
$h = 6$	(0.0)	13.40%	9.93%
	(1.6)	12.07%	

MAPE = mean absolute percentage error

Δ = improvement in forecasting accuracy

Estimated values are computed recursively using dynamic n-step-ahead forecasts with a rolling window of 48 months for each model. The evaluation period is Jan 2008–Oct 2014.

Table 4.4: Nowcasting and forecasting accuracy of the seasonal AR(1) benchmark model (0.0) and the extended models (1.0)–(1.6) that include Google Index 2008–2014.

Allikas: (Tuhkuri, J., Forecasting unemployment with Google Searches, ETLA Working Papers No 35., 2016, [<https://www.etla.fi/wp-content/uploads/ETLA-Working-Papers-35.pdf>] 31.10.2016).

Parendus on välja toodud muutuse protsendina.

SUMMARY

FORECASTING WITH GOOGLE QUERY DATA USING TIME SERIES ESTONIAN EXAMPLES

Margus Matkur

Internet is global. Google web search engine is the most common web search engine in the Internet and it is used to execute billions of queries every day – Big Data. The data about the queries are recorded and are available to everyone via Google Trends. In the present study it is examined whether the collected Google query keyword frequency data are useful in forecasting economic processes using Estonian examples.

In the first chapter of the study an introduction about the interested target groups and the subjects was presented. Followed by the short overview about the previously committed researches. In the second chapter four Estonian economic indicators were introduced: Estonian unemployment rate, the number of people accommodated, retail via mail-order or Internet and the number of the property sales transactions. Next the Google Trends application was introduced and within one-month period of time the raw data was collected every day. Based on observations the Google index (GI) was constructed for each example. An overview of the possible approaches to forecasting was given. In the third chapter of the study the graphical analysis between the Google index and the economic process examples was made. The graphical analysis was the tool to choose the applicable keywords set into the Google Trends observations in the first place. The forecasting models were constructed and the forecasts were performed. The parameter estimation technique called the Prediction Error Method (PEM) was used to optimize the predictive performance of the prediction models.

Based on the “forecasting the past” (*ex post*) method it is confirmed that it is reasonable to bring Google query keyword frequency data along in forecasting Estonian

unemployment rate (extended model MAPE improvement compared to benchmark model MAPE was 1,53%), retail via mail-order or Internet (extended model MAPE improvement compared to benchmark model MAPE was 5,15%) and the number of the property sales transactions (extended model MAPE improvement compared to benchmark model MAPE was 3,37%). In this study the valid set of the keywords was not found to improve the accuracy of the number of people accommodated forecasting (extended model MAPE decrease compared to benchmark model MAPE was 0,24%). Based on the “forecasting the future” (*ex ante*) method the future projections of the time series was composed for each example. In this work, the repeatable algorithm was created of how to improve the accuracy of forecasting based on Internet query data. Compared to the previously committed researches the elimination of the instability of the Google raw data and the prediction models thickening plus the extension of the horizon can be pointed out.

The study looked into the economical subjects in which there is a heightened interest in the wider population or in other words there exists a sufficient data about the keywords connected to the subject in the Google Trends. The results found in this study are consistent with earlier studies and showed that the Google queries keyword frequency data has capability to improve the forecasting accuracy.

The present study and the created algorithm can be used for a software application development that predicts economic indicators in real time. Such a forecasting tool can substantially simplify decision making processes inside commercial or public organizations. Also the study can be useful in more detailed investigation about one particular example also. The author sees the potential for further research of the Estonian unemployment rate, focusing to the suitable forecasting models development. Also the possible link between Google query keyword frequency data and the trading activity in the stock markets is interesting subject to explore.

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, MARGUS MATKUR

(autori nimi)

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose
PROGNOOSIMINE GOOGLE OTSINGU MÄRKŠÖNADE SAGEDUSE ABIL
AEGRIDADE MEETODITEGA

(lõputöö pealkiri)

mille juhendaja on ANDRES VÕRK

(juhendaja nimi)

- 1.1.reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
- 1.2.üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.
3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, **24.05.2017**